

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 9.50981 Seconds
(without alignments)
2953.573 Million cell updates/sec

Title: US-10-650-585-15
Perfect score: 1532
Sequence: 1 AAGLIRACMLVRKAGGHV.....RGVAKAVDFIPVSEMTWR 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1490	97.3	3010	1 A45573	genome polypeptide
2	1473	96.1	3010	1 GNMWVJ	genome polypeptide
3	1463	95.5	3010	1 GNMWTV	genome polypeptide
4	1434	93.6	3010	1 S18030	genome polypeptide
5	1424	93.0	3010	1 GNMWTC	genome polypeptide
6	1360	88.8	3011	1 S40770	genome polypeptide
7	1355	88.4	3011	1 GNMWV3	genome polypeptide
8	1342	87.6	3011	1 GNMWCH	genome polypeptide
9	1192	77.8	3014	1 JCS620	genome polypeptide
10	1130	73.8	3033	1 JQ1303	genome polypeptide
11	1110	72.5	3033	1 GNMWV8	genome polypeptide
12	397.5	25.9	3005	2 T08841	polypeptide - dour
13	341	22.3	2970	2 T08835	polypeptide - marm
14	101	6.6	600	2 B46642	DNA-directed DNA p
15	97.5	6.4	1085	2 T03531	cobn protein homol
16	95.5	6.2	470	2 JCS4098	lecithylase 6-hyd
17	93.5	6.1	353	2 G87392	conserved hypochet
18	93	6.1	660	1 VHMW2	structural protein
19	92.5	6.0	706	2 S33761	transferrin precu
20	92.5	6.0	716	2 G83612	hypothetical prote
21	91	5.9	904	2 A84212	hypothetical prote
22	90.5	5.9	868	2 H81775	aconitate hydratase
23	90	5.9	2796	2 JCS4743	fatty-acid synthase
24	89.5	5.8	7463	2 T36248	CDA peptidase
25	88	5.7	659	2 B44212	structural protein
26	87	5.7	3412	1 GNMWV8	genome polypeptide
27	86.5	5.6	470	1 GNMWV8	genome polypeptide
28	85.5	5.6	348	2 H70549	exo-alpha-sialidas
29					probable pdmb prot

30	85	5.5	470	1 NMIV9	exo-alpha-sialidas
31	85	5.5	707	2 D84154	cadherin-transpor
32	84.5	5.5	1057	2 T18171	hugl protein - hum
33	84	5.5	233	2 T35594	hypothetical prote
34	84	5.5	3069	2 H70656	fatty-acid synthase
35	83.5	5.5	299	2 AH3447	cytochrome-c oxidase
36	83.5	5.5	315	2 AG3361	hypothetical prote
37	83.5	5.5	538	2 S23409	D-alanyl-D-alanine
38	83.5	5.5	990	2 S67499	glutamate synthase
39	83	5.4	398	2 B71284	probable periplasm
40	82.5	5.4	453	1 NMIV3	exo-alpha-sialidas
41	82.5	5.4	868	2 C81200	aconitate hydratase
42	82.5	5.4	5627	2 C83339	hypothetical prote
43	82	5.4	404	2 F87393	amino transferase
44	82	5.4	612	2 T01346	glucan 1,4-alpha-g
45	81.5	5.3	278	2 S51747	light harvesting c

ALIGNMENTS

RESULT 1
A45573 genome polypeptide - hepatitis C virus (strain UT)
N/Contains: capsid protein C; envelope protein M; hepatitis C virus genome from a single Japanese carrier; se
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
C/date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C/accession: A45573
R/Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, N
Virus Res. 23, 39-53, 1992
A/Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier; se
A/reference number: A45573; MUID:92295714; PMID:1316627
A/accession: A45573
A/status: preliminary
A/molecule type: DNA
A/residues: 1-3010 <TAN>
A/cross-references: GB:DI1168; GB:DO1171; NID:9221612; PIDN:BA01943.1; PID:9221613
A/experimental source: HCV-UT
A/Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)
C/Superfamily: hepatitis C virus genome polypeptide
F/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine
F/116-191/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: envelope protein M #status predicted <EMP>
F/192-389/Product: major envelope protein E #status predicted <MEB>
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1007-1615/Product: hepatitis C virus genome polypeptide #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)
F/1312-1317/Region: nucleotide-binding motif B
F/1316-1319/Region: DEXH motif
F/1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match	97.3%	Score 1490;	DB 1;	Length 3010;
Best Local Similarity	96.2%	Pred. No. 9.5e-12;		
Matches 281;	Conservative 5;	Mismatches 6;	Indels 0;	Gaps 0;
QY	1	AAGLIRACMLVRKAGGHVQMAFKMLALGTYYDHLTPLODAHAGLRDLAAVAPV	60	
DB	915	AAGLIRACMLVRKAGGHVQMAFKMLALGTYYDHLTPLODAHAGLRDLAAVAPV	974	
QY	61	ITSDMEVKIITGADTAAAGDIIISGLPVARRGRILLGPADNFGCGKRLIAPITAYSQ	120	
DB	975	VSDMEVKIITGADTAAAGDIIISGLPVARRGRILLGPADNFGCGKRLIAPITAYSQ	1034	
QY	121	QRRGLGCIITSLTRDNNOVGEVQVSTATQSLATCVNGVCTVPHGAGSKTLAPK	180	
DB	1035	QRRGLGCIITSLTRDNNOVGEVQVSTATQSLATCVNGVCTVPHGAGSKTLAPK	1094	
QY	181	GPITQYTNVDDLVGQAPGARSMTPTCGSSDLYLVTRADYIPVRRRDSGSLIS	240	

Db 1095 GPTQMTYNTVDQDLVGMWHPAGARSLPTCTGSSDLYLVRHADVIPVRRGDSRGLLS 1154
QY 241 PRPVSYLKSSGGPILCPSSGHAVGIFRAAVCTRGVAKAVDFIVESMETTR 292
Db 1155 PRPVSYLKSSGGPILCPSSGHAVGIFRAAVCTRGVAKAVDFIVESMETTR 1206

RESULT 2

GNMWTCJ
genome polypeptide - hepatitis C virus (strain J)
N/Contains: capsid protein C; envelope protein E; major envelope protein B; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C/Accession: A39253; PMID:1334449
R/Kato, N.; Hijioka, S.; Ohtsuka, Y.; Nakagawa, M.; Okoshi, S.; Sugimura, T.; Shimoto
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A/Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A/Reference number: A39253; MUID:91088550; PMID:2175903
A/Accession: A39253
A/Molecule type: genomic RNA
A/Residues: 1-3010 <KAT>
A/Cross-references: GB:090308; NID:G221610; PIDN:BA14233.1; PID:G221611
R/Kato, N.; Okoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A/Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
A/Reference number: PS0085
A/Accession: PS0086
A/Molecule type: genomic RNA
A/Residues: 2650-2707 <KAT>
A/Experimental source: Japanese isolate
C/Comment: The cleavage sites of this polypeptide have not been determined.
C/Superfamily: hepatitis C virus genome polypeptide
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: major envelope protein M #status predicted <EMP>
F/132-389/Product: nonstructural protein NS1 #status predicted <NS1>
F/390-729/Product: nonstructural protein NS2 #status predicted <NS2>
F/730-1006/Product: nonstructural protein NS3 #status predicted <NS3>
F/1007-1615/Product: nonstructural protein NS4 #status predicted <NS4>
F/1320-1237/Region: nucleotide-binding motif A (P-loop)
F/1312-1317/Region: nucleotide-binding motif B
F/116-1319/Region: DEXH motif
F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F/196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 96.1%; Score 1473; DB 1; Length 3010;
Best Local Similarity 94.5%; Pred. No. 2,9e-119;
Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHGLRLAVAVEPV 60
Db 915 AAGLIRACMLVRKKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHGLRLAVAVEPV 974
QY 61 IFSDMEVKITITMGADTAACGDIISGLPVARSSEIILGPADNFEQGMRLAPITAYSQ 120
Db 975 VFSDEMEKITITMGADTAACGDIISGLPVARSSEIILGPADNFEQGMRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVEGVVSTATQSFATCVNGVCMWTFHAGSKTLAAGPK 180
Db 1035 QTRGLGCIITSLTGRDNQVEGVVSTATQSFATCVNGVCMWTFHAGSKTLAAGPK 1094
QY 181 GPTQMTYNTVDQDLVGMWHPAGARSLPTCTGSSDLYLVRHADVIPVRRGDSRGLLS 240
Db 1095 GPTQMTYNTVDQDLVGMWHPAGARSLPTCTGSSDLYLVRHADVIPVRRGDSRGLLS 1154
QY 241 PRPVSYLKSSGGPILCPSSGHAVGIFRAAVCTRGVAKAVDFIVESMETTR 292
Db 1155 PRPVSYLKSSGGPILCPSSGHAVGIFRAAVCTRGVAKAVDFIVESMETTR 1206

RESULT 3

GNMWTCJ
genome polypeptide - hepatitis C virus (strain Taiwan)
N/Contains: capsid protein C; envelope protein M; hepatitis A (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
A/Note: host Homo sapiens (man)
C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C/Accession: A40244
R/Chan, P.T.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A/Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A/Reference number: A40244; MUID:92230206; PMID:1334449
A/Accession: A40244
A/Molecule type: genomic RNA
A/Residues: 1-3010 <CHE>
A/Cross-references: GB:M64754
C/Superfamily: hepatitis C virus genome polypeptide
C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F/1-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: major envelope protein M #status predicted <EMP>
F/192-389/Product: nonstructural protein NS1 #status predicted <NS1>
F/390-729/Product: nonstructural protein NS2 #status predicted <NS2>
F/730-1006/Product: nonstructural protein NS3 #status predicted <NS3>
F/1007-1615/Product: nonstructural protein NS4 #status predicted <NS4>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)
F/1312-1317/Region: nucleotide-binding motif B
F/1316-1319/Region: DEXH motif
F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F/196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,

Query Match 95.5%; Score 1463; DB 1; Length 3010;
Best Local Similarity 93.8%; Pred. No. 2.1e-118;
Matches 274; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHGLRLAVAVEPV 60
Db 915 AAGLIRACMLVRKKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHGLRLAVAVEPV 974
QY 61 IFSDMEVKITITMGADTAACGDIISGLPVARSSEIILGPADNFEQGMRLAPITAYSQ 120
Db 975 VFSDEMEKITITMGADTAACGDIISGLPVARSSEIILGPADNFEQGMRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVEGVVSTATQSFATCVNGVCMWTFHAGSKTLAAGPK 180
Db 1035 QTRGLGCIITSLTGRDNQVEGVVSTATQSFATCVNGVCMWTFHAGSKTLAAGPK 1094
QY 181 GPTQMTYNTVDQDLVGMWHPAGARSLPTCTGSSDLYLVRHADVIPVRRGDSRGLLS 240
Db 1095 GPTQMTYNTVDQDLVGMWHPAGARSLPTCTGSSDLYLVRHADVIPVRRGDSRGLLS 1154
QY 241 PRPVSYLKSSGGPILCPSSGHAVGIFRAAVCTRGVAKAVDFIVESMETTR 292
Db 1155 PRPVSYLKSSGGPILCPSSGHAVGIFRAAVCTRGVAKAVDFIVESMETTR 1206

RESULT 4

GNMWTCJ
genome polypeptide - hepatitis C virus (isolate JKI)
N/Contains: capsid protein C; envelope protein M; hepatitis A (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
A/Variety: isolate JKI
C/Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
C/Accession: S18030; S33570; K48333; S18029
R/Honda, M.; Kaneko, S.; Maeshima, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A/Description: A whole genome of hepatitis C virus cDNA was isolated from a single patier
A/Reference number: S18028
A/Accession: S18030
A/Molecule type: genomic RNA

A:Residues: 1-3010 <HON>
A:Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA3793.1; PID:G59479
A:Experimental source: isolate JX1 from an individual
R:Honda, M.; Kaneko, S.; Uncura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
A:Reference number: A48332; MUID:91119270; PMID:8360322
A:Accession: S33570
A:Molecule type: genomic RNA
A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>
A:Cross-references: EMBL:X61591
A:Note: this sequence is inconsistent with the nucleotide translation
A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
as Trp, and TTC for residue 771 as Ser
A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:P121748)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:116-191/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:132-389/Product: major envelope protein E #status predicted <MEB>
F:330-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:112-1317/Region: nucleotide-binding motif B
F:116-1319/Region: DEXH motif
F:116-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 93.6%; Score 1434; DB 1; Length 3010;
Best Local Similarity 92.8%; Pred. No. 7e-116;
Matches 271; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHVQMAFMKLAALGTIVYDHLTPLOMAHAGRLDAVAPEV 60
Db 915 AAGLIRACMLVRKAGGHVQMAFMKLAALGTIVYDHLTPLOMAHAGRLDAVAPEV 974

QY 61 IFSDMEVKIITWGADTAACGDIISGLPVARSRGREILGPADNFGQGRLLAPITAYSQ 120
Db 975 VFSDMETKIIITWGADTAACGDIISGLPVARSRGREILGPADNFGQGRLLAPITAYSQ 1034

QY 121 QTRGLGCIITSLTGRDNQGEVQVSTATQSLATCVNGVCTVTHGAGSKTLAEPK 180
Db 1035 QTRGLGCIITSLTGRDNQGEVQVSTATQSLATCVNGVCTVTHGAGSKTLAEPK 1094

QY 181 GPITQYTNVDQDLVGMQAPPARBSMTPTCTGSSDLVLTFRADVTPRRRDSGSLLS 240
Db 1095 GPITQYTNVDQDLVGMQAPPARBSMTPTCTGSSDLVLTFRADVTPRRRDSGSLLS 1154

QY 241 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETMR 292
Db 1155 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETMR 1206

RESULT 5
GENWTC
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitisin (BC 3.4.21.98) (nonstructu
protein NS4b; nonstructural protein NS4d; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: A38465
R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;
J. Virol. 65, 1105-1113, 1991
A:Title: Structure and organization of the hepatitis C virus genome isolated from human
A:Reference number: A38465; MUID:91140696; PMID:1847440
A:Accession: A38465
A:Molecule type: genomic RNA
A:Residues: 1-3010 <TAK>
A:Cross-references: EMBL:M58335; NID:G329770; PIDN:AAA72945.1; PID:G329771
C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F:116-191/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:132-389/Product: major envelope protein E #status predicted <MEB>
F:330-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:112-1317/Region: nucleotide-binding motif B
F:116-1319/Region: DEXH motif
F:116-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 93.0%; Score 1424; DB 1; Length 3010;
Best Local Similarity 92.1%; Pred. No. 5.2e-115;
Matches 269; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHVQMAFMKLAALGTIVYDHLTPLOMAHAGRLDAVAPEV 60
Db 915 AAGLIRACMLVRKAGGHVQMAFMKLAALGTIVYDHLTPLOMAHAGRLDAVAPEV 974

QY 61 IFSDMEVKIITWGADTAACGDIISGLPVARSRGREILGPADNFGQGRLLAPITAYSQ 120
Db 975 VFSDMETKIIITWGADTAACGDIISGLPVARSRGREILGPADNFGQGRLLAPITAYSQ 1034

QY 121 QTRGLGCIITSLTGRDNQGEVQVSTATQSLATCVNGVCTVTHGAGSKTLAEPK 180
Db 1035 QTRGLGCIITSLTGRDNQGEVQVSTATQSLATCVNGVCTVTHGAGSKTLAEPK 1094

QY 181 GPITQYTNVDQDLVGMQAPPARBSMTPTCTGSSDLVLTFRADVTPRRRDSGSLLS 240
Db 1095 GPITQYTNVDQDLVGMQAPPARBSMTPTCTGSSDLVLTFRADVTPRRRDSGSLLS 1154

QY 241 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETMR 292
Db 1155 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETMR 1206

RESULT 6
S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitisin (BC 3.4.21.98) (nonstructu
protein NS4b; nonstructural protein NS4d; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:G221586; PIDN:BA01582.1; PID:G221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsunoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:G221511; PIDN:BA00705.1; PID:G221512
A:Experimental source: isolate HC-U1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:116-191/Product: capsid protein C #status predicted <CPC>
F:132-389/Product: major envelope protein E #status predicted <MEB>
F:330-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1316-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 88.8%; Score 1360; DB 1; Length 3011;

Best Local Similarity 85.6%; Pred. No. 1.9e-109; Mismatches 20; Indels 0; Gaps 0;

Matches 249; Conservative 22; Mismatches 20; Indels 0; Gaps 0;

QY 2 QGIRACMLVRKAGHYVOMAFKMLALGTYYVDHLPLODMAHGRLDLAVAVEPVI 61
DB 916 QGIRACMLVRKAGHYVOMAFKMLALGTYYVDHLPLODMAHGRLDLAVAVEPVI 975
QY 62 FSDMEVKITITGADTAACGDIISGLPVASARGREILLGPADNFEQGMRLAPITAYSOQ 121
DB 976 FSDMEVKITITGADTAACGDIISGLPVASARGREILLGPADNFEQGMRLAPITAYSOQ 1035
QY 122 TRGLIGCIITSLTGRDNQVGEVQVYSTAQSLATCINGVCMVTHGAGRTITASPFG 181
DB 1036 TRGLIGCIITSLTGRDNQVGEVQVYSTAQSLATCINGVCMVTHGAGRTITASPFG 1095
QY 182 PITOMYTNVDDPLVGMOPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRGRSGSLSP 241
DB 1096 PITOMYTNVDDPLVGMOPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRGRSGSLSP 1155
QY 242 RPVSYLKSGSGGPIICPSGHAAGIFRAAVCTRGVAKAVDFPVESEMTTR 292
DB 1156 RPVSYLKSGSGGPIICPSGHAAGIFRAAVCTRGVAKAVDFPVESEMTTR 1206

RESULT 7

GNMW3
genome polypeptide - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A39166; P00404
R:Chao, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A:Title: Genetic organization and diversity of the hepatitis C virus.
A:Reference number: A39166; MUID:91172826; PMID:1848704
A:Accession: A39166
A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-references: GB:M67321; NID:G329873; PIDN:AAA45676.1; PID:G329874
R:Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A:Reference number: P00393; MUID:92268871; PMID:1316939
A:Accession: P00403
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-references: DDBJ:D10128
A:Experimental source: Isolates E-b16
A:Accession: P00404
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: Isolates E-b17
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolase; nonstructura
F:1-115/Product: capsid protein C #status predicted <CPC>
F:115-191/Product: envelope protein M #status predicted <EPM>
F:191-389/Product: major envelope protein E #status predicted <ME>
F:389-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:729-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis virus #status predicted <NS3>
F:1615-1862/Product: nucleotide-binding motif A (P-loop)
F:1862-2013/Product: nonstructural protein NS4a #status predicted <N4a>
F:2014-3011/Product: nonstructural protein NS4b #status predicted <N4b>
F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196-209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 88.4%; Score 1355; DB 1; Length 3011;

Best Local Similarity 85.2%; Pred. No. 5.2e-109; Mismatches 20; Indels 0; Gaps 0;

Matches 248; Conservative 23; Mismatches 20; Indels 0; Gaps 0;

QY 2 QGIRACMLVRKAGHYVOMAFKMLALGTYYVDHLPLODMAHGRLDLAVAVEPVI 61
DB 916 QGIRACMLVRKAGHYVOMAFKMLALGTYYVDHLPLODMAHGRLDLAVAVEPVI 975
QY 62 FSDMEVKITITGADTAACGDIISGLPVASARGREILLGPADNFEQGMRLAPITAYSOQ 121
DB 976 FSDMEVKITITGADTAACGDIISGLPVASARGREILLGPADNFEQGMRLAPITAYSOQ 1035
QY 122 TRGLIGCIITSLTGRDNQVGEVQVYSTAQSLATCINGVCMVTHGAGRTITASPFG 181
DB 1036 TRGLIGCIITSLTGRDNQVGEVQVYSTAQSLATCINGVCMVTHGAGRTITASPFG 1095
QY 182 PITOMYTNVDDPLVGMOPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRGRSGSLSP 241
DB 1096 PITOMYTNVDDPLVGMOPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRGRSGSLSP 1155
QY 242 RPVSYLKSGSGGPIICPSGHAAGIFRAAVCTRGVAKAVDFPVESEMTTR 292
DB 1156 RPVSYLKSGSGGPIICPSGHAAGIFRAAVCTRGVAKAVDFPVESEMTTR 1206

RESULT 8

GNMW3
genome polypeptide - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchausti, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738
R:Inchausti, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F:1-115/Product: capsid protein C #status predicted <CPC>
F:115-191/Product: envelope protein M #status predicted <EPM>
F:191-389/Product: major envelope protein E #status predicted <ME>
F:389-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:729-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis virus #status predicted <NS3>
F:1615-1862/Product: nucleotide-binding motif A (P-loop)
F:1862-2013/Product: nonstructural protein NS4a #status predicted <N4a>
F:2014-3011/Product: nonstructural protein NS4b #status predicted <N4b>
F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
F:2014-3011/Product: nonstructural protein NS4b #status predicted <N4b>
F:1316-1319/Region: DEXH motif
F:1316-1319/Region: DEXH motif
Query Match 87.6%; Score 1342; DB 1; Length 3011;
Best Local Similarity 84.5%; Pred. No. 7e-108; Mismatches 25; Indels 0; Gaps 0;
Matches 246; Conservative 25; Mismatches 20; Indels 0; Gaps 0;
QY 2 QGIRACMLVRKAGHYVOMAFKMLALGTYYVDHLPLODMAHGRLDLAVAVEPVI 61

Db 916 QGLIRICALARKINGHYVQMAIKLGAITGTCTVYNNLAPLBDPAHNGLDLAAVAEPV 975
Qy 62 FSDNEVKIITWGADTAACGDIISGLPVARSRGREILIGPADNFGCGWRLAPITAYSQ 121
Db 976 FSRRETKLITWGADTAACGDIISGLPVARSRGREILIGPADNFGCGWRLAPITAYSQ 1035
Qy 122 TRGLIGCIITSLTRDKNQVEGEVQVSTATQSFATCVNGVCTVPHGAGSKTLAPGK 181
Db 1036 TRGLIGCIITSLTRDKNQVEGEVQVSTATQSFATCVNGVCTVPHGAGSKTLAPGK 1095
Qy 182 PITQMTYNTVDQDLVGWQAPPGARSMTPTCGSSDLVYTRHADYIPVRRGDSRGSILSP 241
Db 1096 PVICTYNTVDQDLVGWQAPPGARSMTPTCGSSDLVYTRHADYIPVRRGDSRGSILSP 1155
Qy 242 RPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 292
Db 1156 RPISYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 1206

RESULT 9
JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
C:Species: hepatitis C virus
C:Keywords: ATP; glycoprotein; hydrolyase; polypeptide binding; P-loop; polypeptide; transme
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: J05620
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
A:Reference number: J05620; MUID:97366593; PMID:9223423
A:Accession: J05620
A:Molecule type: mRNA
A:Residues: 1-3014 <CHA>
A:Cross-references: GB:Y13184
A:Experimental source: genotype 5a, which predominates in South Africa
A:Note: the translation of the nucleotide sequence is not complete in this paper
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolyase; polypeptide binding; P-loop; polypeptide; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:384-408/Region: hypervariable #status predicted
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F:1008-1616/Product: nonstructural protein NS2 #status predicted <NS3>
F:1231-1238/Region: nucleotide-binding motif A (P-loop)
F:1313-1318/Region: nucleotide-binding motif A (P-loop)
F:1317-1320/Region: DEXH motif
F:1317-1863/Product: nonstructural protein NS4 #status predicted <N4>
F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4b>
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2210-2249/Region: interferon sensitivity determining #status predicted

Query March 77.8%; Score 1192; DB 1; Length 3014;
Best Local Similarity 72.6%; Pred. No. 7.9e-95;
Matches 212; Conservative 41; Mismatches 39; Indels 0; Gaps 0;
Qy 1 AGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYDHTPLQDMAHAGRLDAVAEPV 60
Db 916 ARALLRMLCTLAKHLVYKQVQALHLGRLTGYIYDHLAPMDMAASGRELTVATEPI 975
Qy 61 IFSDEVKIITWGADTAACGDIISGLPVARSRGREILIGPADNFGCGWRLAPITAYSQ 120
Db 976 VFSDEVKIITWGADTAACGDIISGLPVARSRGREILIGPADNFGCGWRLAPITAYSQ 1035
Qy 121 QTRGLIGCIITSLTRDKNQVEGEVQVSTATQSFATCVNGVCTVPHGAGSKTLAPGK 180
Db 1036 QTRGLIGCIITSLTRDKNQVEGEVQVSTATQSFATCVNGVCTVPHGAGSKTLAPGK 1095
Qy 181 GPITQMTYNTVDQDLVGWQAPPGARSMTPTCGSSDLVYTRHADYIPVRRGDSRGSILS 240

Db 1096 GPVYQMTYNTVDQDLVGWQAPPGARSMTPTCGSSADLVYTRHADYIPARRRGTBASILS 1155
Qy 241 RPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 292
Db 1156 RPISYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 1207

RESULT 10
J01303
genome polyprotein - hepatitis C virus (isolate HC-16)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
C:Species: hepatitis C virus
C:Keywords: ATP; glycoprotein; hydrolyase; P-loop; polypeptide; transme
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C:Accession: J01303
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurat, K.; Itzuka, H.; Machida, A.; Miyakawa, Y.;
J. Gen. Virol. 72, 2697-2704, 1991
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a hum
A:Reference number: J01303; MUID:92044440; PMID:1658196
A:Accession: J01303
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D00944; NID:9221650; PTDN:BA00792.1; PID:9221651
A:Experimental source: isolate HC-16 from a Japanese individual
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolyase; P-loop; polypeptide; serine proteinase; transme
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepacivirin #status predicted <NS3>
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4 #status predicted <N4>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,477,534,545,558,578,627,649,1091,1217,1259,2038,283

Query March 73.8%; Score 1130; DB 1; Length 3033;
Best Local Similarity 69.2%; Pred. No. 1.9e-89;
Matches 202; Conservative 42; Mismatches 46; Indels 0; Gaps 0;

Qy 1 AGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYDHTPLQDMAHAGRLDAVAEPV 60
Db 916 ARALLRMLCTLAKHLVYKQVQALHLGRLTGYIYDHTPLQDMAHAGRLDAVAEPV 978
Qy 61 IFSDEVKIITWGADTAACGDIISGLPVARSRGREILIGPADNFGCGWRLAPITAYSQ 120
Db 976 IFSDEVKIITWGADTAACGDIISGLPVARSRGREILIGPADNFGCGWRLAPITAYSQ 1038
Qy 121 QTRGLIGCIITSLTRDKNQVEGEVQVSTATQSFATCVNGVCTVPHGAGSKTLAPGK 180
Db 1039 QTRGLIGCIITSLTRDKNQVEGEVQVSTATQSFATCVNGVCTVPHGAGSKTLAPGK 1098
Qy 181 GPITQMTYNTVDQDLVGWQAPPGARSMTPTCGSSDLVYTRHADYIPVRRRDSRGSILS 240
Db 1099 GPITQMTYNTVDQDLVGWQAPPGARSMTPTCGSSDLVYTRHADYIPVRRRDSRGSILS 1158
Qy 241 RPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 292
Db 1159 RPISYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 1210

RESULT 11
GNV08
genome polyprotein - hepatitis C virus (strain HC-J8)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
C:Species: hepatitis C virus
C:Keywords: ATP; glycoprotein; hydrolyase; P-loop; polypeptide; transme
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40250; F00357; F00559
R:Okamoto, H.; Kurat, K.; Okada, S.; Yamamoto, K.; Itzuka, H.; Taraka, T.; Fukuda, S.;

Virology 188, 331-341, 1992
 A>Title: Full length sequence of a hepatitis C virus genome having poor homology to repo
 A/Reference number: A40250; MUID:92230232; PMID:1114459
 A/Accession: A40250
 A/Molecule type: genomic RNA
 A/Residues: 1-3033 <OK>
 A/Cross-references: GB:D10988; GB:D01221; NID:9221608; PIDN:BA01761.1; PID:9221609
 R/Chan, S.M.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
 J. Gen. Virol. 73, 1131-1141, 1992
 A>Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
 A/Reference number: PQ0393; MUID:9226871; PMID:116939
 A/Accession: PQ0397
 A/Molecule type: genomic RNA
 A/Residues: 2678-2754 <CHA>
 A/Cross-references: DDBJ:D10134
 A/Experimental source: isolate E-312
 R/Kato, N.; Ootsuyama, Y.; Okoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohnc
 Biochem. Biophys. Res. Commun. 181, 279-285, 1991
 A>Title: Distribution of plural HCV types in Japan.
 A/Reference number: PQ0554; MUID:92068204; PMID:1720309
 A/Accession: PQ0559
 A/Molecule type: mRNA
 A/Residues: 2678-2729 <KAT>
 A/Cross-references: GB:D10562; GB:D09518; NID:9221523; PIDN:BA01418.1; PID:9221524
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1.115/Product: capsid protein C #status predicted <CPC>
 F:1.16/Product: envelope protein M #status predicted <EMP>
 F:1.16-191/Product: major envelope protein E #status predicted <N3>
 F:330-733/Product: nonstructural protein NS1 #status predicted <N1>
 F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1011-1619/Product: hepatitis virus #status predicted <NS3>
 F:1334-1241/Region: nucleotide-binding motif A (P-loop)
 F:1316-1321/Region: nucleotide-binding motif B
 F:1320-1323/Region: DEH motif
 F:1520-1866/Product: nonstructural protein NS4 #status predicted <N4>
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23
 { Query Match 72.5%; Score 1111; DB 1; Length 3033;
 Best Local Similarity 67.8%; Pred. No. 8,7e-88;
 Matches 198; Conservative 44; Mismatches 50; Indels 0; Gaps 0;
 QY 1 AAGLIRACMLVRKAAGHYVOMAFKALITGVYDHLTPLODMAHAGIRDLAVVEPY 60
 DB 919 AHALIRVCTIVKILAGARYIOMLITIGRMTGYIYDHSPLSTMAAOGIRDLAVVEPY 978
 QY 61 IFSDMEKXITTWGADTACGDIISGLPVSARGREITLGPADNFEQGMELLAPITAYSQ 120
 DB 979 VFSEMEKXIVTWGAEIVACGDIHLGLPVSARLGRVLLGPADGTSKGMKLLAPITAYQ 1038
 QY 121 QTRGLACITITSLGRDKNOVEGVYVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 180
 DB 1039 QTRGLALIVSLTGRDKNOAGQVLSVTOFTLOTSISGLMTVYHAGKNTLACPK 1098
 QY 181 GPTQXTNVDDQIVGQAPGARSMTPCTCGSSDLYLTRHADVIVRRRGRSGSLIS 240
 DB 1099 GPTQXTNVSAEGDVLWGAPSPGTSLDPCCTGAAVDLTVTRADVIVRRKXDRGSLIS 1158
 QY 241 PRPVSYLKSSGSPILCPSGHVAIGIFRAAVCTRGVAAVDFIVESMETTR 292
 DB 1159 PRPLSTLKSSGSPVILCSRGHVAIGLFRAAVCAVAAASIDFIVESGLDVATR 1210
 RESULT 12
 T08841
 polyprotein - douroucouli hepatitis GB virus A
 C:Species: douroucouli hepatitis GB virus A
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
 C/Accession: T08841
 R/Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
 J. Gen. Virol. 79, 41-45, 1998

A>Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
 A/Reference number: Z16486; MUID:98120818; PMID:9460920
 A/Accession: T08841
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-3005 <EKK>
 A/Cross-references: EMBL:AF023425; NID:92828599; PIDN:AA040501.1; PID:92828600
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: polyprotein

Query Match 25.9%; Score 397.5; DB 2; Length 3005;
 Best Local Similarity 34.8%; Pred. No. 9.4e-26;
 Matches 95; Conservative 46; Mismatches 119; Indels 13; Gaps 5;
 QY 23 AFMKLAALITGVYVDHLTPLODMAHAGIRDLAVVEPYIFSDMEKXITTWGADTACGDI 82
 DB 887 AFVRLERSGTGLTHOQGVXGAAILXDLGVALLVEVSTAYRDCYIVRDAARTLACGR 946
 QY 83 ISGLPVSARRRREILLG--PADNFEQGMELLAPITAYSQOTRLLCIIITSLGRDKNO 140
 DB 947 VEGLPVARRDDEVLYGVFPSPRALPGFVFTAVVV--MORGLEFSSVNTSKMRDRE 1005
 QY 141 VEGEVQVSTATQSEFLATCVNGVCMVTFHAGSKTLAEPKPIQMTNVDDQIVGQAP 200
 DB 1006 HEGSLVUGITSTSMGTCVNGVMTTFHGSNARTLAPVGPVCRKMSDSDVAVYPLP 1065
 QY 201 FGARSMTCTCGSSDLYIVTRHADVIVRRRGRSGSLISFRPVSYLKSSGSPILCPG 260
 DB 1066 SGASGLEBCKCGTOSVWCIRN--DGALCHGSLXVLEIDLPTEISDRSGSSPILCDEG 1123
 QY 261 HAVGIFRAAVCTRGV-----AKAVDFIPVES 286
 DB 1124 HAVGIM-VSLHKGVTGVRVYKFWETLPKDS 1155

RESULT 13
 T08839
 polyprotein - marmoset hepatitis GB virus A
 C:Species: marmoset hepatitis GB virus A
 C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
 C/Accession: T08839
 R/Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
 J. Gen. Virol. 79, 41-45, 1998
 A>Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
 A/Reference number: Z16486; MUID:98120818; PMID:9460920
 A/Accession: T08839
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: genomic RNA
 A/Residues: 1-2970 <EKK>
 A/Cross-references: EMBL:AF023424; NID:92828597; PIDN:AA040501.1; PID:92828598
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: polyprotein

Query Match 22.3%; Score 341; DB 2; Length 2970;
 Best Local Similarity 30.5%; Pred. No. 7.6e-21;
 Matches 85; Conservative 46; Mismatches 96; Indels 52; Gaps 9;
 QY 45 WAHAG-----LRDLAVVEPYIFSDMEKXITTWGADTACGDIISGLPVSARRGRE 95
 DB 891 YAHAGQVTRRAEQLRQMGFLPEVAVHPEBCAVRDAARTLSCGSGVHGPVAAARGDE 950
 QY 96 ILLGPADNFEQGMEL-----LAPITAYSQOTRLLCIIITSLGRDKNOVEGVYVS 149
 DB 951 VILGVLNV---WELPGFVFTAVVVH--HKGGFQGVVTSMTGMDDETHAVNVVVLG 1005
 QY 150 TATOSFLATCVNGVCMVTFHAGSKTLAEPKPIQMTNVDDQIVGQAPGARSMTPC 209
 DB 1006 TSTTRSMGTCVNGVMTTFHGSNARTLAPVGPVCRKMSDSDVAVYPLPVGAKCLEPC 1065
 QY 210 TCGSSDLYIVTRHADVIVRRRGRSGSLIS-----PPVSYLKSSGSPILCP 258
 DB 1066 KQPGGQWV-----RND--GALCHGTLGRVLEIDLPAFLCDFRSGSSGSPILCD 1112

QY 259 SGHAGVGFRAAVCTRG-----VAKAVDFIPVESMETT 290
 DB 1113 EGHAVGVL-ISTLHRSRVGTGIRYTKPMTLPREATHIT 1150

RESULT 14

B46642
 DNA-directed DNA polymerase (EC 2.7.7.7) alpha/DNA primase (EC 2.7.7.-) complex 68k chaf
 C:Species: Mus musculus (house mouse)
 C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
 C:Accession: B46642
 R: Miyazawa, H.; Itzumi, M.; Tada, S.; Takada, R.; Masutani, M.; Ui, M.; Hanaoka, F.
 J. Biol. Chem. 268, 8111-8122, 1993
 A>Title: Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase al
 A:Reference number: A46642, MUID:93216788; PMID:8463324
 A:Accession: B46642
 A>Status: preliminary
 A:Molecule type: mRNA; protein
 A:Residues: 1-600 <MTY>
 A:Cross-references: GB:D13546; NID:g303658; PIDN:BAA02746.1; PID:g303659
 A:Experimental source: FM3A cells
 A>Note: sequence extracted from NCBI backbone (NCBIN:129148, NCBI:P.129149)
 C:Keywords: nucleotidyltransferase

Query Match 6.6%; Score 101; DB 2; Length 600;
 Best Local Similarity 24.8%; Pred. No. 0.77;
 Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

QY 4 LIRACMLVRKAGGHVQM-AFMKLAIT-----GTYVDH-----TPLODMA 46
 DB 27 LAELCVLYRQTEDEWVSELIAFCTAGTKCLTVDLINSFEYEVINKKLAKAMHSASKDSG 86
 QY 47 HAGRDLAVAVEPIFSDMEVKITWGADTAACDI--ISGLP-----VSARRGREI 96
 DB 87 HAGRDIT-VSIQELIEAEEETLLSYTTPSKGPLKRVSTPEPTPLTKRSVARSPPQ- 144
 QY 97 LLGPADNFEQGKMLLAITVYSQQTRGLGCIITSLTGRDKNQVEGEVQVSTATQSF 156
 DB 145 LLSPTS-----FSPSATPSQK-----YTSRTNR-----GEVVTTFGSAQ--- 178
 QY 157 ATCNNGVCMVTFHAGSKTL--AGPKGPIQMTYTNVDQDLVG 196
 DB 179 -----GLWSGRGSGSVSLKVGDPDEPLTGSYKAMFQQLMG 215

RESULT 15

T03531
 COB protein homolog - Rhodobacter capsulatus

C:Species: Rhodobacter capsulatus
 C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 18-Feb-2000
 C:Accession: T03531
 R: Vilek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
 Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
 A>Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003
 A:Reference number: Z14955; MUID:97404404; PMID:9256491
 A:Accession: T03531
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1085 <VLC>
 A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAC16184.1; PID:g3128332
 C:Genetics:
 A:Map position: 1
 C:Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase

Query Match 6.4%; Score 97.5; DB 2; Length 1085;
 Best Local Similarity 22.2%; Pred. No. 3.2;
 Matches 82; Conservative 30; Mismatches 116; Indels 141; Gaps 18;

QY 14 AAGGTYVGMAMFKLAALGTYYDHLTP-----IQDMAHAGRLDLAVAVEPIFSDMEV 67
 DB 144 AGGAVAAQAALAQALAGLTA---PVPQDKRLPDM--GYDPAAGVTPAPEGDGFL 196
 QY 68 KIITW-----GADTAACGDIISGLPVSARRGREI---LLGPADNFEQGKMLLAIT- 115

DB 197 ALVSFYSVLTADAPIDALIHGL-----RARGFAAGLFAFSLKSEGLGALQARDG 251
 QY 116 -----TAYSQQ-----TRGL-----LSCI 129
 DB 252 RAPALVVMATAFSAAGADGTPPDYDVPVQVALSTARKDMALADRLSPADLAMHV 311
 QY 130 ITSLTGR-----DKQVEGEVQV-----STATQS 154
 DB 312 LPEVDRLFAGVYSFKAPRPDLEFSKFAHRPDPAVEALDRICGHRARQTPADR 371
 QY 155 FLATCVN--GVCWTVFHAGSKTLAGPKGPIQMTYTNVDQDLVGMQADPGARSMTPCTC 211
 DB 372 KLAIVLSTYFGRSQGMHAHVGDLALASTEALAQALATE-----GVAIAPGALSTLTTA 425
 QY 212 ----GSSDLXYTRHADVIPIARRGDSRGLSPRPVSYLKSGSGGPLLCPGHAVGIFR 267
 DB 426 RLTPPLSD-YLTA-LDTLPEKLRAD-----LTPAMGPPEADPA-CADGAPR 468
 QY 268 AAVCTEGVA 276
 DB 469 PALARRGAA 477

Search completed: May 6, 2004, 09:37:21
 Job time: 10.5098 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 6.09603 Seconds
(without alignments)
2494.160 Million cell updates/sec

Title: US-10-650-585-15

Perfect score: 1532
Sequence: 1 AAGLIRACMLVRRKAGSHV.....RGVAKAVDPFVPSMETTNR 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1490	97.3	3010	1	POLG_HCVUT
2	1473	96.1	3010	1	POLG_HCVTA
3	1463	95.5	3010	1	POLG_HCVTW
4	1424	93.0	3010	1	POLG_HCVXK
5	1355	88.4	3011	1	POLG_HCVL
6	1342	87.6	3011	1	POLG_HCVI
7	1130	73.8	3033	1	POLG_HCVU6
8	1111	72.5	3033	1	POLG_HCVU8
9	101	6.6	600	1	POLG_HCVU8
10	93	6.1	660	1	POLG_HCVU8
11	93	6.1	660	1	POLG_HCVU8
12	92.5	6.0	660	1	POLG_HCVU8
13	89.5	5.8	659	1	POLG_HCVU8
14	87	5.7	3414	1	POLG_HCVU8
15	87	5.7	3412	1	POLG_HCVU8
16	87	5.7	3414	1	POLG_HCVU8
17	86.5	5.6	470	1	POLG_HCVU8
18	85	5.5	470	1	POLG_HCVU8
19	85	5.5	470	1	POLG_HCVU8
20	84.5	5.5	1705	1	POLG_HCVU8
21	84	5.5	309	1	POLG_HCVU8
22	84	5.5	339	1	POLG_HCVU8
23	84	5.5	470	1	POLG_HCVU8
24	83.5	5.5	538	1	POLG_HCVU8
25	83.5	5.5	538	1	POLG_HCVU8
26	82.5	5.4	453	1	POLG_HCVU8
27	82	5.4	309	1	POLG_HCVU8
28	82	5.4	403	1	POLG_HCVU8
29	82	5.4	482	1	POLG_HCVU8
30	82	5.4	612	1	POLG_HCVU8
31	81.5	5.3	398	1	POLG_HCVU8
32	81.5	5.3	1022	1	POLG_HCVU8
33	81.5	5.3	1022	1	POLG_HCVU8

RESULT 1	ID	Sequence	Standard	PRT	3010 AA
POLG_HCVUT	1	1490	97.3	3010	1
POLG_HCVUT	1	1473	96.1	3010	1
POLG_HCVUT	1	1463	95.5	3010	1
POLG_HCVUT	1	1424	93.0	3010	1
POLG_HCVUT	1	1355	88.4	3011	1
POLG_HCVUT	1	1342	87.6	3011	1
POLG_HCVUT	1	1130	73.8	3033	1
POLG_HCVUT	1	1111	72.5	3033	1
POLG_HCVUT	1	101	6.6	600	1
POLG_HCVUT	1	93	6.1	660	1
POLG_HCVUT	1	93	6.1	660	1
POLG_HCVUT	1	92.5	6.0	659	1
POLG_HCVUT	1	89.5	5.8	659	1
POLG_HCVUT	1	87	5.7	3414	1
POLG_HCVUT	1	87	5.7	3412	1
POLG_HCVUT	1	87	5.7	3414	1
POLG_HCVUT	1	86.5	5.6	470	1
POLG_HCVUT	1	85	5.5	470	1
POLG_HCVUT	1	85	5.5	470	1
POLG_HCVUT	1	84.5	5.5	1705	1
POLG_HCVUT	1	84	5.5	309	1
POLG_HCVUT	1	84	5.5	339	1
POLG_HCVUT	1	84	5.5	470	1
POLG_HCVUT	1	83.5	5.5	538	1
POLG_HCVUT	1	83.5	5.5	538	1
POLG_HCVUT	1	82.5	5.4	453	1
POLG_HCVUT	1	82	5.4	309	1
POLG_HCVUT	1	82	5.4	403	1
POLG_HCVUT	1	82	5.4	482	1
POLG_HCVUT	1	82	5.4	612	1
POLG_HCVUT	1	81.5	5.3	398	1
POLG_HCVUT	1	81.5	5.3	1022	1
POLG_HCVUT	1	81.5	5.3	1022	1

ALIGNMENTS

RESULT 1	ID	Sequence	Standard	PRT	3010 AA
POLG_HCVUT	1	1490	97.3	3010	1
POLG_HCVUT	1	1473	96.1	3010	1
POLG_HCVUT	1	1463	95.5	3010	1
POLG_HCVUT	1	1424	93.0	3010	1
POLG_HCVUT	1	1355	88.4	3011	1
POLG_HCVUT	1	1342	87.6	3011	1
POLG_HCVUT	1	1130	73.8	3033	1
POLG_HCVUT	1	1111	72.5	3033	1
POLG_HCVUT	1	101	6.6	600	1
POLG_HCVUT	1	93	6.1	660	1
POLG_HCVUT	1	93	6.1	660	1
POLG_HCVUT	1	92.5	6.0	659	1
POLG_HCVUT	1	89.5	5.8	659	1
POLG_HCVUT	1	87	5.7	3414	1
POLG_HCVUT	1	87	5.7	3412	1
POLG_HCVUT	1	87	5.7	3414	1
POLG_HCVUT	1	86.5	5.6	470	1
POLG_HCVUT	1	85	5.5	470	1
POLG_HCVUT	1	85	5.5	470	1
POLG_HCVUT	1	84.5	5.5	1705	1
POLG_HCVUT	1	84	5.5	309	1
POLG_HCVUT	1	84	5.5	339	1
POLG_HCVUT	1	84	5.5	470	1
POLG_HCVUT	1	83.5	5.5	538	1
POLG_HCVUT	1	83.5	5.5	538	1
POLG_HCVUT	1	82.5	5.4	453	1
POLG_HCVUT	1	82	5.4	309	1
POLG_HCVUT	1	82	5.4	403	1
POLG_HCVUT	1	82	5.4	482	1
POLG_HCVUT	1	82	5.4	612	1
POLG_HCVUT	1	81.5	5.3	398	1
POLG_HCVUT	1	81.5	5.3	1022	1
POLG_HCVUT	1	81.5	5.3	1022	1

```

DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5a.
DR InterPro: IPR001650; HCV_RdRp.
DR InterPro: IPR004109; Peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSV1r.
DR Pfam: PF01543; HCV_core; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR SMART: SMO0487; DEXDC; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2529 2529
FT CARBOHYD 2788 2788
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C7435D642BB CRC64;
Query Match 97.3%; Score 1490; DB 1; Length 3010;
Best Local Similarity 96.2%; Pred. No. 6; 7e-122;
Matches 281; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
1 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTYYDHTLPLODMAHAGLDLAFAVEPV 60

```

```

Db 915 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTYYDHTLPLODMAHAGLDLAFAVEPV 974
Qy 61 IFSDMEVITITMGADTAACDIIISGLPYASARGREIILGPDNDNEGOWRLAPITAYSQ 120
Db 975 VFSDEMTIITMGADTAACDIIISGLPYASARGREIILGPDNDNEGOWRLAPITAYSQ 1034
Qy 121 QTRGLAGCIITSLTRGRDNQVEGEVQVSTATQSFATCVNGVCWTYFPHGASKTLAQP 180
Db 1035 QTRGLAGCIITSLTRGRDNQVEGEVQVSTATQSFATCVNGVCWTYFPHGASKTLAQP 1094
Qy 181 GPITOMTNNVDIVGMAPPGASMPCTCGSSDVIYMRHATVIVRRGDSGLS 240
Db 1095 GPITOMTNNVDIVGMAPPGASMPCTCGSSDVIYMRHATVIVRRGDSGLS 1154
Qy 241 PRPVSYLNGSSGPPILCPSGHVGIFRAAVCTRGVAVAVDFIPVESMTTNR 292
Db 1155 PRPVSYLNGSSGPPILCPSGHVGIFRAAVCTRGVAVAVDFIPVESMTTNR 1206

RESULT 2
POLG_HCVUA STANDARD; PRT; 3010 AA.
AC 26662;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polypeptide (Coltains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OC Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11116;
RN [1]
RP MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RT Proc. Natl. Acad. Sci. U.S.A. 87:9524-9526(1990).
RL [2]
RP MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RT Proc. Natl. Acad. Sci. U.S.A. 87:9524-9526(1990).
RL [2]
DISCUSSION OF SEQUENCE.
RP MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RT FEBS Lett. 280:325-328(1991).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

CC ----- EMBL: D90208; BAA14233.1; -
 DR PIR: A39253; GNMVCT.
 DR HSP: P26663; IUXP.
 DR MEROPS; S29.001; -
 DR InterPro: IPR009003; Cys_ser_crypsin.
 DR InterPro: IPR004410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Replidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR Pfam: PF0186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KV Transmembrane; Nonstructural protein; Hydrophobic; Serine protease.
 FT INT1 MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT SITE 1083 1083
 FT ACT SITE 1107 1107
 FT ACT SITE 1165 1165
 FT NP BIND 1230 1237
 FT SITE 1316 1319
 FT CAROHD 196 196
 FT CAROHD 209 209
 FT CAROHD 234 234
 FT CAROHD 250 250
 FT CAROHD 305 305
 FT CAROHD 417 417
 FT CAROHD 423 423
 FT CAROHD 448 448
 FT CAROHD 532 532
 FT CAROHD 556 556
 FT CAROHD 576 576
 FT CAROHD 623 623
 FT CAROHD 645 645
 FT CAROHD 2041 2041
 FT CAROHD 2077 2077
 FT CAROHD 2240 2240
 FT CAROHD 2768 2768
 FT CAROHD 3010 AA; 327017 MM; AA939794F46DB185 CRC64;
 SQ SEQUENCE

Query Match 96.1%; Score 1473; DB 1; Length 3010;
 Best Local Similarity 94.5%; Pred. No. 2,1e-120;
 Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;
 QY 1 AGLIRACMLVRKAGGHHYVOMAPFKALALGTYYVHDLTPLODPAHAGRLDAVAPEV 60
 DB AGLIRACMLVRKAGGHHYVOMAPFKALALGTYYVHDLTPLODPAHAGRLDAVAPEV 974
 QY 61 IPSDEVKIITWGAADTAACGDIISGLPVASARGREIILGPADNFGCGWRLIAPITAYSQ 120
 DB VFSDEKTLITWGAADTAACGDIISGLPVASARGREIILGPADNFGCGWRLIAPITAYSQ 1034
 QY 121 QTRGLIGCIITSITGRDNKQVEGVVSTATQSLATCNGVCMVTHGASKTLAPK 180
 DB QTRGLIGCIITSITGRDNKQVEGVVSTATQSLATCNGVCMVTHGASKTLAPK 1094
 QY 181 GPTQMTYVDDPLVGMQAPGARSGMPTCTGSSDLYIVTRHADVIPYRRRDSRGLIS 240
 DB GPTQMTYVDDPLVGMQAPGARSGMPTCTGSSDLYIVTRHADVIPYRRRDSRGLIS 1154
 QY 241 PPEVSTLKSGSGGPIICPSGHAAGIFRAVCTRGYAKAVDFIVESMETTR 292
 DB PPEVSTLKSGSGGPIICPSGHAAGIFRAVCTRGYAKAVDFIVESMETTR 1206
 RESULT 3
 POLG HCVTM STANDARD; PRT; 3010 AA.
 AC P29846;
 ID POLG_HCVTM
 AD 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [containing: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 DE Hepatitis C virus (isolate Taiwan) (HCV).
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9220206; PubMed=1314449;
 RA Chen P.U., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' terminus of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113 (1992).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 hydrophobic, suggesting a possible membrane-related function. NS3
 and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
 (RNA) (N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

ID	NAME	SEQUENCE	START	END	SCORE	DB	LENGTH	GAPS	INDELS
FT	CARBOHYD	2240	2240						
FT	CARBOHYD	2529	2529						
FT	CARBOHYD	2788	2788						
SO	SEQUENCE	3010 AA; 327047 MW; AAD267B5C0FE215 CRC64;							
Query Match			95.5%	Score 1463; DB 1;			Length 3010;		
Best Local Similarity			93.8%	Ident. No. 1.5e-119;					
Matches 274;			Conservative	9;	Mismatches	0;	Indels	0;	Gaps 0;
QY	1	AGGLIRACMLYKAKAGHYVQVAFKLAALGTYYDHLTLPLODAHAHGLRLAVAVPV	60						
DB	915	AGGLIRACMLYKAKAGHYVQVAFKLAALGTYYDHLTLPLODAHAHGLRLAVAVPV	974						
QY	61	IFSDMEVKIITWGADTAAAGDIIISGLPVSAARGREILLGPADNFEQGWRLIAPITAYSQ	120						
DB	975	VFSDMETKIITWGADTAAAGDIIISGLPVSAARGREILLGPADNFEQGWRLIAPITAYSQ	1034						
QY	121	OTRGLGLGIIITSLGRDKNOYEGEVQVSTNOSGLACNVGCMCTVTHGSGKTLAPK	180						
DB	1035	OTRGLGLGIIITSLGRDKNOYEGEVQVSTNOSGLACNVGCMCTVTHGSGKTLAPK	1094						
QY	181	GPITOMYTNVDQLVGMQAPPARSMTPECTGSSDPLVLRHADVIPIRRRGDSRGSLLS	240						
DB	1095	GPITOMYTNVDQLVGMQAPPARSMTPECTGSSDPLVLRHADVIPIRRRGDSRGSLLS	1155						
QY	241	PRPVSYLKSGSGGELPCSGHANGIFRAACTRTGYAKAVDIPVPSMETTKR	292						
DB	1155	PRPVSYLKSGSGGELPCSGHANGIFRAACTRTGYAKAVDIPVPSMETTKR	1206						
RESULT 4									
POLG_HCVX									
ID	POLG_HCVX	STANDARD;	PRT;	3010 AA.					
AC	P26663;								
DT	01-AUG-1992 (Rel. 23, Created)								
DT	01-AUG-1992 (Rel. 23, Last sequence update)								
DE	10-OCT-2003 (Rel. 42, Last annotation update)								
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);								
DE	Envelope glycoprotein E1 (GP23) (GP25); Envelope glycoprotein E2								
DE	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)								
DE	(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)								
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein								
DE	NS4B (P56) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].								
OS	Hepatitis C virus (Isolate BK) (HCV).								
CC	Virusess; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;								
OC	Hepacivirus.								
OX	NCBI_TaxID=11105;								
NP	[1]								
RP	SEQUENCE FROM N.A.								
RP	MEDLINE=91140698; PubMed=18474440;								
RX	Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,								
RA	Onishi E., Andoh T., Yoshida I., Okayama H.;								
RT	"Structure and organization of the hepatitis C virus genome isolated								
RT	from human carriers."								
RL	J. Virol. 65:1105-1113(1991).								
RN	[2]								
RP	SEQUENCE OF 1487-1500.								
RP	MEDLINE=96235224; PubMed=8647104;								
RX	Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;								
RT	"Nonstructural protein 3 of hepatitis C virus inhibits								
RT	phosphorylation mediated by cAMP-dependent protein kinase."								
RL	Eur. J. Biochem. 237:611-618(1996).								
RN	[3]								

RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinduebler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of HX strain hepatitis C
 RT virus: a 2.2 Å resolution structure in a hexagonal crystal form";
 RL Protein Sci. 7:837-847(1998).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polypeptide, commonly with Asp or Glu in the P6
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
 CC (RNA) (N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and RNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-eb.ch/announce/>
 CC or send an email to license@isb-eb.ch).
 CC -----
 DR EMBL: M58335; AAA72945.1; -;
 DR PIR: A38465; GAWVTC.
 DR PDB: 1A10; 25-MAR-98.
 DR PDB: 1UXE; 14-JAN-98.
 DR PDB: 1NS3; 08-APR-98.
 DR PDB: 1C2P; 15-NOV-00.
 DR PDB: 1GSJ; 08-NOV-99.
 DR PDB: 1GX5; 09-APR-02.
 DR PDB: 1QIV; 26-JUN-00.
 DR PDB: 8OHM; 20-APR-99.
 DR MEROPS: S29.001; -;
 DR MEROPS: U39.001; -;
 DR InterPro: IPR009003; Cys_ser_cryptin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4A.
 DR InterPro: IPR001490; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR004109; peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4A; 1.
 DR Pfam: PF01001; HCV_NS4B; 1.
 DR Pfam: PF01506; HCV_NS5A; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.

FT CHAIN	1	115	CAPSID PROTEIN C (POTENTIAL).
FT CHAIN	116	191	MATRIX PROTEIN (POTENTIAL).
FT CHAIN	192	333	MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN	384	729	NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN	730	1006	NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN	1007	1615	PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN	1616	1862	NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN	1863	2013	NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN	2014	3010	RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM	347	369	POTENTIAL.
FT ACT_SITE	1083	1083	CHARGE RELAY SYSTEM.
FT ACT_SITE	1107	1107	CHARGE RELAY SYSTEM.
FT ACT_SITE	1165	1165	CHARGE RELAY SYSTEM.
FT NE_BIND	1230	1237	ATP (POTENTIAL).
FT SITE	1316	1319	DECH BOX.
FT CARBOHYD	196	196	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	209	209	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	234	234	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	305	305	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	417	417	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	423	423	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	430	430	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	448	448	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	532	532	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	540	540	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	556	556	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	623	623	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	645	645	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	2041	2041	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	2077	2077	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	2240	2240	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	2529	2529	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	2788	2788	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1031	1035	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX	1039	1047	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1050	1050	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1059	1063	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1066	1074	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1075	1076	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1077	1081	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX	1082	1085	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1086	1087	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1090	1092	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1093	1094	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1095	1097	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1101	1103	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1104	1107	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1108	1112	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1120	1120	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1122	1132	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1129	1133	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1135	1136	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1139	1144	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1149	1157	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX	1158	1161	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1162	1163	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1165	1166	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1168	1171	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1172	1174	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1175	1186	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1187	1188	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1189	1197	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX	1198	1202	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1203	1204	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1680	1688	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE	3010 AA;	327189 MM;	F8422D5ECCDFD9C CRC64;

Query Match 93.0%; Score 1424; DB 1; Length 3010;
 Best local Similarity 92.1%; Pred. No. 4e-116;
 Matches 269; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

Query Match 88.4%; Score 1355; DB 1; Length 3011;
 Best Local Similarity 85.2%; Pred. No. 4.4e-110;
 Matches 248; Conservative 23; Mismatches 20; Indels 0; Gaps 0;

CC 2 QGLIRACMLYKKAAGHYVQMAFMKALATGTYTDTLPLQWANAAGLDLAVANPVI 61
 CC |||||
 CC 916 QGLIRFCALARKIMGHVYQVWVIRIKGALGTIVYVNHLPRLPWAHAGLDLAVANPVI 975
 CC |||||
 CC 62 FSDMEVLTITWGADTAACGDISGLPVASRGREIILGPADNFEQGMRLAPITAVSQ 121
 CC |||||
 CC 976 FSGMETLTITMGADTAACGDIINLPASARRRILILGPADGWSKMLLAPITAAQ 1035
 CC |||||
 CC 122 TRGLGCTITSLTRGRDKNQVEGVQVSTATQSLATCVNGCVTWVHGAGSKTLAPKG 181
 CC |||||
 CC 1036 TRGLGCTITSLTRGRDKNQVEGVQVSTAAQTFLATCINGCVTWVHGAGRTIIASPKG 1095
 CC |||||
 CC 182 PITGMVTVNODIVGMQAPGASMTPTCTGSSDLVYTHADVTPRRRGDSRLSP 241
 CC |||||
 CC 1096 PVIMVTVNDDLVGMVPAQSRSLPTCTGSSDLVYTHADVTPRRRGDSRLSP 1155
 CC |||||

CC 242 RPVSYLKSSGCGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETNR 292
 CC |||||
 CC 1156 RPIYLKSSGCGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVENMETNR 1206
 CC |||||

RESULT 6
 POLG HCVH STANDARD; PRT; 3011 AA.

AC P27958;
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-AUG-1992 (Rel. 42, Last annotation update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Genome polypeptide (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4) (Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).
 OS Hepatitis C virus (isolate H) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_Taxid=11106;
 CC [1]

RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92052256; PubMed=1658800;
 RX Inchauspe G.; Zebedee S.; Lee D.H.H.; Sugitani M.; Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates."
 RT Proc. Natl. Acad. Sci. U.S.A. 86:10292-10296(1991).
 RL [2]
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N.; Hesson T.; Cable M.; Hong Z.; Kwong A.D.; Le H.V.; Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain."
 RT Nat. Struct. Biol. 4:463-467(1997).
 RL [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L.; Morgenstern K.A.; Griffith J.P.; Dwyer M.D.; Thomson J.A.,
 RA Mucke M.A.; Lin C.; Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding."
 RL Structure 6:99-100(1998).
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN

CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
 CC (RNA) (N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M67463; AAA4534.1; -
 CC PIR: A36814; GNVVCH
 CC PDB: 1HEI; 25-NOV-98.
 CC PDB: 1AIV; 16-FEB-99.
 CC PDB: 1AIR; 17-JUN-98.
 CC MEROPS: S29.001; -
 CC MEROPS: U39.001; -
 CC TRANSFAC: T04155; -
 CC InterPro: IPR009003; Cys_Set_trypsin.
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR001650; Helicase_C.
 CC InterPro: IPR004109; Peptidase_C29.
 CC InterPro: IPR007095; RNA_pol_ds_PS.
 CC InterPro: IPR007094; RNA_pol_PstI.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00271; helicase_C; 1.
 CC Pfam: PF00398; Viral_RdRP; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00487; DEXDc1; 1.
 CC K0 Polypeptin; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;
 CC 3D-structure.
 CC INIT MET 1 1
 CC CHAIN 1 191
 CC CHAIN 192 383
 CC CHAIN 384 746
 CC CHAIN 747 809
 CC CHAIN 810 1026
 CC CHAIN 1027 1657
 CC CHAIN 1658 1711
 CC CHAIN 1712 1972
 CC CHAIN 1973 2420

REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 CAPSID PROTEIN C.
 ENVELOPE GLYCOPROTEIN E1.
 ENVELOPE GLYCOPROTEIN E2.
 PROTEIN P7.
 NONSTRUCTURAL PROTEIN NS2.
 PROTEASE/HELICASE NS3.
 NONSTRUCTURAL PROTEIN NS4A.
 NONSTRUCTURAL PROTEIN NS4B.
 NONSTRUCTURAL PROTEIN NS5A.

FT	HELIX	1606	1611
FT	TURN	1614	1618
FT	STRAND	1622	1623
FT	STRAND	1627	1627
FT	STRAND	1635	1636
FT	HELIX	1640	1652
EQ	SEQUENCE	3011 AA; 327142 MW; 772CBB29C9CD4753 CRC64;	
Query Match			
Best Local Similarity 87.6%; Score 1342; DB 1; Length 3011;			
Matches 246; Conservative 25; Pident. No. 6,1e-109; Mismatches 20; Indels 0; Gaps 0;			
QY	2	QGLIRACMLVRKAAGGYVQMAFMKLAALGCTVYDHLLTFLQWMAHGLADLVAVEPI	61
Db	916	QGLIRICALARKAAGHYVMAIKKALGTCVYNLAFLRMAHGLDLVAVEPV	975
QY	62	FSMEVEKILIMWCAADTAACDIIISGLPSASARGREILLGPADNEGGQMRLLAPITAYSOQ	121
Db	976	FSMRKELIMWADTAACDIIINGLPVSARGREILLGPADGNVSKMRLLAPITAYAOQ	1035
QY	122	TRGLIGCTITSLGRDRNQVEGEVQVSTATQSEFLATCNGVCWTYFHGSGKTLAPKG	181
Db	1036	TRGLIGCTITSLGRDRNQVEGEVQVSTATQSEFLATCNGVCWTYFHGSGKTLAPKG	1095
QY	182	PIQGMATNVQDLYGMQAPRGASMPCTCGSSPLVYVTHADVIPIVRKRGDSRLSP	241
Db	1096	PIQGMATNVQDLYGMQAPRGASMPCTCGSSPLVYVTHADVIPIVRKRGDSRLSP	1155
QY	242	RPVSYLKGSGGPGILCPSGHAGVIFRAAVCTRGVAKADEFIPVESMETIWR	292
Db	1156	RPVSYLKGSGGPGILCPGTGHAAGVIFRAAVCTRGVAKADEFIPVENLETTWR	1206
RESULT 7			
POLG_HCVU6		STANDARD;	PRT; 3033 AA.
AC	P26660;		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Genome polyprotein [Contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin) (EC 3.4.21.98); Nonstructural protein NS4 (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.46)].		
OS	Hepatitis C virus (isolate HC-J6) (HCV).		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.		
CC	NCBI_TaxId=11113;		
CC	[1]		
RP	SEQUENCE FROM N.A.		
CC	MEDLINE=J2044440; PubMed=1658196;		
CC	Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H., Machida A., Miyakawa Y., Mayumi M.;		
RA	"Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions."		
RT	J. Gen. Virol. 72:2697-2704(1991).		
RL	-1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.		
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position. Cys or Thr in P1 and Ser or Ala in P1'.		
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N).		
CC	-1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.		
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.		

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: D13546; BAA02746.1; -
 DR PIR: B46642; B46642.
 DR MGI: 99690; Polaz.
 DR InterPro: IPR007200; DNA_pol_alpha.B.
 DR Pfam: PF04058; DNA_pol_alpha.B; 1.
 KM DNA replication; Nuclear protein; Phosphorylation.
 FT DOMAIN 101 107 POLY-GLU.
 FT DOMAIN 115 157 PRO/SER/THR-RICH (HYDROPHILIC).
 SQ SEQUENCE 600 AA; 66267 MW; 79F94BEEFE33FEEB CRC64;

Query Match 6.6%; Score 101; DB 1; Length 600;
 Best Local Similarity 24.8%; Pred. No. 0.36;
 Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

QY 4 LIBACMLVRYKAGGHVYQW-AFMELALT-----GTYVDHL-----TPLODMA 46
 DB 27 LAELCVLYRQEDMGSELNFTSAGKCLTVDILNSFEYVLNKKLSAMHSASKDSG 86
 QY 47 HAGLRDLAVAPVIFSDMEVKITWGADTAAAGDI--ISGLP-----VSARRGREI 96
 DB 87 HAGTRDI-VSIQELIEAEEBEETLLSYTPSKPLKRVSTPETPLTKSVARSAPRC- 144
 QY 97 LUGADNFEQGGKMLAPITVYSGQTGLGCLITSLTGKDKVNEEYVQVSTANQSL 156
 DB 145 LLSPPS-----FSPSATPSQK-----YTSRTNR-----GEVYTRGSAQ-- 178
 QY 157 ATCVNGVCMVTFHAGSKTL--AGPKQPIQMTYNTVQDVLG 196
 DB 179 -----GLSMGSGSGSVSLKVVDPEPLTSGSYKAMQQLMG 215

RESULT 10

VST2 HEVBU STANDARD; PRT; 660 AA.
 ID VST2 HEVBU
 AC P23326;
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Structural protein 2 precursor (ORF2).
 OS Hepatitis E virus (strain Burma) (HEV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage;
 OC Hepatitis E-like viruses.
 OX NCBI_TaxID=31767;

RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=92024067; PubMed=1926770;
 RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
 RA Fry K.E., Reyes G.R.;
 RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
 RT full-length viral genome";
 RL Virology 185:120-131(1991).

CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M73318; AAA45736.1; -
 DR PIR: C40778; VHWKH2.
 DR InterPro: IPR004261; SP2.
 DR InterPro: IPR008975; Viral_cap_coat.

DR Pfam: PF03014; SP2; 1.
 KM Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 660 STRUCTURAL PROTEIN 2.
 SQ SEQUENCE 660 AA; 70978 MW; 5832A013CCCA61C CRC64;

Query Match 6.1%; Score 93; DB 1; Length 660;
 Best Local Similarity 19.0%; Pred. No. 2;
 Matches 72; Conservative 47; Mismatches 119; Indels 140; Gaps 16;

QY 1 AGLIRACMLVRYKAGGHVYQWAFMKLAALGTYYVDHLTPLODMAHAGLRDLAVAPV 60
 DB 188 AATATRYRDLVNVNAGVAISISFMPQTTPTSV-----DNNSITSTVRLVDPG 239
 QY 61 IFSDMEVKITWGADTAAAGDIISGLPVARRGREILLGPAD--NEEGGWELAPI-TA 117
 DB 240 IASELYI-----PSERLHYRQGRSVETSGVA 267
 QY 118 YSQQTRGLL-----GCITSLTG----- 135
 DB 268 EEBATSGILVWLCIHGSLVSYNTPTYGALGLDPALEFRLNTPGNTNTRVSRYSYA 327
 QY 136 --RDKVQGEVQVSTANQSLA---TCVNGV-----CWTVEH----- 169
 DB 328 RHLRERGADTAEITTTATRFMKDLYFTSTNGVEIGRIALLTLFNADTLGLPIEL 387
 QY 170 --GAG-----SKTLGPKG-PITOMYNTVDODLVGMQAPGARSMTPCTCGSSDLYLV-- 219
 DB 388 ISAGGGLFYSRPVVSANSEPTVKLYTSVENA-----QDQKGLAIPHIDILGSRVYIQY 443
 QY 220 -TRHADVIVRRRGDSRG-SLSPRPVSLK-----GSSGGLPLCSGHAVGIF 266
 DB 444 DNOHEQDRPTPPAPSRPFSVLRANDVWLSTLAAYDOSTYSGSSGTPYV--SDSVTLV 501
 QY 267 RAAVCTRGVAKAVDFIPV 284
 DB 502 NVATGQAVARSLDWTKV 519

RESULT 11

VST2 HEVPA STANDARD; PRT; 660 AA.
 ID VST2 HEVPA
 AC P33436;
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Structural protein 2 precursor (ORF2).
 OS Hepatitis E virus (strain Pakistan) (HEV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage;
 OC Hepatitis E-like viruses.
 OX NCBI_TaxID=33774;

RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=92115700; PubMed=1713327;
 RA Tearev S.A., Emerson S.U., Reyes G.R., Teareva T.S., Legters L.J.,
 RA Malik I.A., Iqbal M., Purcell R.H.;
 RT "Characterization of a prototype strain of hepatitis E virus";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).

CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M80581; AAA45727.1; -
 DR PIR: C40778; VHWKH2.
 DR InterPro: IPR004261; SP2.
 DR InterPro: IPR008975; Viral_cap_coat.

DR Pfam; PF03014; SP2; 1.
 KM Signal.
 FT SIGNAL 1 22 BY SIMILARITY.
 FT CHAIN 23 660 STRUCTURAL PROTEIN 2.
 SQ SEQUENCE 660 AA; 70980 MW; 8085BC53CFB46FD3 CRC64;
 Query Match 6.1%; Score 93; DB 1; Length 660;
 Best Local Similarity 19.0%; Pred. No. 2;
 Matches 72; Conservative 49; Mismatches 117; Indels 140; Gaps 17;

QY 1 AGLIRACMLYKRAAGHYVQAMFKALALTYVYHLLPLDMMAGLRDLAVAPY 60
 188 ARATIRRPPLVENVAGVAISISFWPQTTPPTVS-----DNKSTSTVRLIVQPG 239
 QY 61 IFSDMEVYITWAGDTAACGDIISGLVPSARAGREILLGPAD--NEFGQMRLL----- 112
 240 IASLVL-----PSEHLHYNQMRVETSGVA 267
 QY 113 -----APITAYSQOT-RGLIGCI-----ITSITGDKXQ----- 140
 268 EEEATSGVLMLCINGSPVNSYTPYTGALGLDFALELEFRNLTPGNTNTRVSRYS97A 327
 QY 141 -----VEGEVQVSTATQSLA-----TCWNGV-----CMTVFF----- 169
 328 RHRRLRGADTAEITTAATREMDLYFTSTNGVETGRGIALTLFPLADTLIGLPTEL 387
 QY 170 --GAG---SKTLAGEKQ-PITQMTYNVDVLVQAPFGASMTPTCGSSDLVY-- 219
 388 ISSAGGQFYSPVANSNGEPTVLYTSVENA-----QODKGIAPHDIDGESRVVYQDY 443
 QY 220 -TRADVIPIRRRDSGC-SLSPRPVSYLK-----SSGGSPILCPGSHAVGIF 266
 444 DNQHEQRPFPSPAPSPSPSVLRANDVLMSTLAETDQSTYSSSTGFPVY--SDSVTLV 501
 QY 267 RAAVCTRGVAKAVDFIPV 284
 502 NVATGAQAVARSLDMTKV 519
 DB

RESULT 12
 TRFE HORSE
 ID TREE HORSE STANDARD; PRT; 706 AA.
 AC P27425;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin).
 GN TF.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 CX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93277958; PubMed=8504171;
 RA Carpenter M.A., Broad T.E.;
 RT "The cDNA sequence of horse transferrin."
 RU Biochim. Biophys. Acta 1173:230-232(1993).
 RP [2]
 RC SEQUENCE FROM N.A.
 RC TISSUE=Extraembryonic tissue;
 RA McDowell K.J., Adams M.H., Baker C.B.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Transferrins are iron binding transport proteins which can bind two atoms of ferric iron in association with the binding of an anion, usually bicarbonate. It is responsible for the transport of iron from sites of absorption and heme degradation to those of storage and utilization. Serum transferrin may also have a further role in stimulating cell proliferation.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.

CC -1- DOMAIN: Composed of two homologous domains.
 CC -1- SIMILARITY: Belongs to the transferrin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M69020; AAA30958.1; -.
 CC EMBL: U21127; AAA63684.1; -.
 CC PIR: S33761; S33761.
 CC HSSP: P02787; IABE.
 CC InterPro: IPR001156; Transferrin.
 CC Pfam: PF00405; Transferrin; 2.
 CC PRINTS: PR00422; TRANSFERRIN.
 CC SMART: SM00094; TR_FBR; 2.
 CC PROSITE: PS00205; TRANSFERRIN_1; 2.
 CC PROSITE: PS00206; TRANSFERRIN_2; 2.
 CC PROSITE: PS00207; TRANSFERRIN_3; 2.
 CC Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KW Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 706 BY SIMILARITY.
 FT REPEAT 20 357 SEROTRANSFERRIN.
 FT REPEAT 358 706 1.
 FT DISULFID 26 64 BY SIMILARITY.
 FT DISULFID 36 55 BY SIMILARITY.
 FT DISULFID 134 215 BY SIMILARITY.
 FT DISULFID 174 190 BY SIMILARITY.
 FT DISULFID 177 198 BY SIMILARITY.
 FT DISULFID 187 200 BY SIMILARITY.
 FT DISULFID 248 262 BY SIMILARITY.
 FT DISULFID 360 623 BY SIMILARITY.
 FT DISULFID 366 398 BY SIMILARITY.
 FT DISULFID 376 399 BY SIMILARITY.
 FT DISULFID 423 701 BY SIMILARITY.
 FT DISULFID 441 664 BY SIMILARITY.
 FT DISULFID 474 550 BY SIMILARITY.
 FT DISULFID 498 692 BY SIMILARITY.
 FT DISULFID 508 522 BY SIMILARITY.
 FT DISULFID 519 533 BY SIMILARITY.
 FT DISULFID 590 604 BY SIMILARITY.
 FT DISULFID 642 647 BY SIMILARITY.
 FT METAL 79 79 IRON 1 (BY SIMILARITY).
 FT METAL 111 111 IRON 1 (BY SIMILARITY).
 FT METAL 209 209 IRON 1 (BY SIMILARITY).
 FT METAL 270 270 IRON 1 (BY SIMILARITY).
 FT METAL 413 413 IRON 2 (BY SIMILARITY).
 FT METAL 449 449 IRON 2 (BY SIMILARITY).
 FT METAL 544 544 IRON 2 (BY SIMILARITY).
 FT METAL 612 612 IRON 2 (BY SIMILARITY).
 FT BINDING 136 136 CARBONATE 1 (BY SIMILARITY).
 FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
 FT BINDING 142 142 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
 FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
 FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 480 480 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
 FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
 FT CARBOHYD 515 515 N-LINKED (GLCNAc...) (POTENTIAL).
 SQ SEQUENCE 706 AA; 78094 MW; 1A0FA566C0409D8A CRC64;

Query Match 6.0%; Score 92.5; DB 1; Length 706;
 Best Local Similarity 21.5%; Pred. No. 2.5;
 Matches 64; Conservative 43; Mismatches 109; Indels 81; Gaps 17;

hydrophobic, suggesting a possible membrane-related function.
NS3 and NS5 may play a role in the viral RNA replication.
-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polypeptide, commonly with Asp or Glu in the p6 position, Cys or Thr in p1 and Ser or Ala in p1'.
-1- CATALYTIC ACTIVITY: N nucleoside triphosphate + (RNA) (N).
-1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and RNA.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U27495; AAA66870.1; --
PIR: A31052; GNMWE.
DR PDB: 1SV3; 10-JUN-96.
DR MEROPS; S07.0PW; --
DR InterPro: IPR009003; Cys_Ser_cryptin.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001122; Flavi_capsidC.
DR InterPro: IPR000335; Flavi_glycoprote.
DR InterPro: IPR000063; Flavi_N.
DR InterPro: IPR001157; Flavi_NS1.
DR InterPro: IPR000752; Flavi_NS2A.
DR InterPro: IPR000487; Flavi_NS2B.
DR InterPro: IPR000404; Flavi_NS4A.
DR InterPro: IPR001528; Flavi_NS4B.
DR InterPro: IPR000208; Flavi_NS5.
DR InterPro: IPR002535; Flavi_propep.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007110; 19-11ke.
DR InterPro: IPR001850; Peptidase_S7.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR InterPro: IPR002877; Rmt_FtsJ.
DR Pfam; PF01003; Flavi_capsid; 1.
DR Pfam; PF02832; Flavi_glycop_C; 1.
DR Pfam; PF00869; Flavi_glycoprot; 1.
DR Pfam; PF00949; Flavi_helicase; 1.
DR Pfam; PF01004; Flavi_M; 1.
DR Pfam; PF00948; Flavi_NS1; 1.
DR Pfam; PF01005; Flavi_NS2A; 1.
DR Pfam; PF01002; Flavi_NS2B; 1.
DR Pfam; PF01350; Flavi_NS4A; 1.
DR Pfam; PF01349; Flavi_NS4B; 1.
DR Pfam; PF00972; Flavi_NS5; 1.
DR Pfam; PF01570; Flavi_propep; 1.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF00271; helicase_C; 1.
DR PRODOM; PD001556; Flavi_glycoprote; 1.
DR PRODOM; PD001496; Flavi_NS1; 1.
DR SMART; SMO0487; DEXDC; 1.
DR SMART; SMO0490; HELICG; 1.
DR PROSITE; PS00690; DEAD_ATP_HELICASE; FALSE NEG.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Hydrolyase; Helicase;
KW ATP-binding; Transmembrane; Nonstructural protein; 3D-structure.
INTL_MET 1

CHAIN 1 112
PROPEP 113 205
CHAIN 206 280
CHAIN 281 776
CHAIN 777 1128
CHAIN 1129 1358
CHAIN 1359 1489

ENVELOPE GLYCOPROTEIN M.
MAJOR ENVELOPE PROTEIN E.
NONSTRUCTURAL PROTEIN NS1.
NONSTRUCTURAL PROTEIN NS2A.
NONSTRUCTURAL PROTEIN NS2B.
NONSTRUCTURAL PROTEIN NS2B.

CHAIN 1490 2110
CHAIN 2110 2259
CHAIN 2259 2360
CHAIN 2360 2511
CHAIN 2511 2612
CHAIN 2612 2713
CHAIN 2713 2814
CHAIN 2814 2915
CHAIN 2915 3016
CHAIN 3016 3117
CHAIN 3117 3218
CHAIN 3218 3319
CHAIN 3319 3420
CHAIN 3420 3521
CHAIN 3521 3622
CHAIN 3622 3723
CHAIN 3723 3824
CHAIN 3824 3925
CHAIN 3925 4026
CHAIN 4026 4127
CHAIN 4127 4228
CHAIN 4228 4329
CHAIN 4329 4430
CHAIN 4430 4531
CHAIN 4531 4632
CHAIN 4632 4733
CHAIN 4733 4834
CHAIN 4834 4935
CHAIN 4935 5036
CHAIN 5036 5137
CHAIN 5137 5238
CHAIN 5238 5339
CHAIN 5339 5440
CHAIN 5440 5541
CHAIN 5541 5642
CHAIN 5642 5743
CHAIN 5743 5844
CHAIN 5844 5945
CHAIN 5945 6046
CHAIN 6046 6147
CHAIN 6147 6248
CHAIN 6248 6349
CHAIN 6349 6450
CHAIN 6450 6551
CHAIN 6551 6652
CHAIN 6652 6753
CHAIN 6753 6854
CHAIN 6854 6955
CHAIN 6955 7056
CHAIN 7056 7157
CHAIN 7157 7258
CHAIN 7258 7359
CHAIN 7359 7460
CHAIN 7460 7561
CHAIN 7561 7662
CHAIN 7662 7763
CHAIN 7763 7864
CHAIN 7864 7965
CHAIN 7965 8066
CHAIN 8066 8167
CHAIN 8167 8268
CHAIN 8268 8369
CHAIN 8369 8470
CHAIN 8470 8571
CHAIN 8571 8672
CHAIN 8672 8773
CHAIN 8773 8874
CHAIN 8874 8975
CHAIN 8975 9076
CHAIN 9076 9177
CHAIN 9177 9278
CHAIN 9278 9379
CHAIN 9379 9480
CHAIN 9480 9581
CHAIN 9581 9682
CHAIN 9682 9783
CHAIN 9783 9884
CHAIN 9884 9985
CHAIN 9985 10086
CHAIN 10086 10187
CHAIN 10187 10288
CHAIN 10288 10389
CHAIN 10389 10490
CHAIN 10490 10591
CHAIN 10591 10692
CHAIN 10692 10793
CHAIN 10793 10894
CHAIN 10894 10995
CHAIN 10995 11096
CHAIN 11096 11197
CHAIN 11197 11298
CHAIN 11298 11399
CHAIN 11399 11400
CHAIN 11400 11501
CHAIN 11501 11602
CHAIN 11602 11703
CHAIN 11703 11804
CHAIN 11804 11905
CHAIN 11905 12006
CHAIN 12006 12107
CHAIN 12107 12208
CHAIN 12208 12309
CHAIN 12309 12410
CHAIN 12410 12511
CHAIN 12511 12612
CHAIN 12612 12713
CHAIN 12713 12814
CHAIN 12814 12915
CHAIN 12915 13016
CHAIN 13016 13117
CHAIN 13117 13218
CHAIN 13218 13319
CHAIN 13319 13420
CHAIN 13420 13521
CHAIN 13521 13622
CHAIN 13622 13723
CHAIN 13723 13824
CHAIN 13824 13925
CHAIN 13925 14026
CHAIN 14026 14127
CHAIN 14127 14228
CHAIN 14228 14329
CHAIN 14329 14430
CHAIN 14430 14531
CHAIN 14531 14632
CHAIN 14632 14733
CHAIN 14733 14834
CHAIN 14834 14935
CHAIN 14935 15036
CHAIN 15036 15137
CHAIN 15137 15238
CHAIN 15238 15339
CHAIN 15339 15440
CHAIN 15440 15541
CHAIN 15541 15642
CHAIN 15642 15743
CHAIN 15743 15844
CHAIN 15844 15945
CHAIN 15945 16046
CHAIN 16046 16147
CHAIN 16147 16248
CHAIN 16248 16349
CHAIN 16349 16450
CHAIN 16450 16551
CHAIN 16551 16652
CHAIN 16652 16753
CHAIN 16753 16854
CHAIN 16854 16955
CHAIN 16955 17056
CHAIN 17056 17157
CHAIN 17157 17258
CHAIN 17258 17359
CHAIN 17359 17460
CHAIN 17460 17561
CHAIN 17561 17662
CHAIN 17662 17763
CHAIN 17763 17864
CHAIN 17864 17965
CHAIN 17965 18066
CHAIN 18066 18167
CHAIN 18167 18268
CHAIN 18268 18369
CHAIN 18369 18470
CHAIN 18470 18571
CHAIN 18571 18672
CHAIN 18672 18773
CHAIN 18773 18874
CHAIN 18874 18975
CHAIN 18975 19076
CHAIN 19076 19177
CHAIN 19177 19278
CHAIN 19278 19379
CHAIN 19379 19480
CHAIN 19480 19581
CHAIN 19581 19682
CHAIN 19682 19783
CHAIN 19783 19884
CHAIN 19884 19985
CHAIN 19985 20086
CHAIN 20086 20187
CHAIN 20187 20288
CHAIN 20288 20389
CHAIN 20389 20490
CHAIN 20490 20591
CHAIN 20591 20692
CHAIN 20692 20793
CHAIN 20793 20894
CHAIN 20894 20995
CHAIN 20995 21096
CHAIN 21096 21197
CHAIN 21197 21298
CHAIN 21298 21399
CHAIN 21399 21400
CHAIN 21400 21501
CHAIN 21501 21602
CHAIN 21602 21703
CHAIN 21703 21804
CHAIN 21804 21905
CHAIN 21905 22006
CHAIN 22006 22107
CHAIN 22107 22208
CHAIN 22208 22309
CHAIN 22309 22410
CHAIN 22410 22511
CHAIN 22511 22612
CHAIN 22612 22713
CHAIN 22713 22814
CHAIN 22814 22915
CHAIN 22915 23016
CHAIN 23016 23117
CHAIN 23117 23218
CHAIN 23218 23319
CHAIN 23319 23420
CHAIN 23420 23521
CHAIN 23521 23622
CHAIN 23622 23723
CHAIN 23723 23824
CHAIN 23824 23925
CHAIN 23925 24026
CHAIN 24026 24127
CHAIN 24127 24228
CHAIN 24228 24329
CHAIN 24329 24430
CHAIN 24430 24531
CHAIN 24531 24632
CHAIN 24632 24733
CHAIN 24733 24834
CHAIN 24834 24935
CHAIN 24935 25036
CHAIN 25036 25137
CHAIN 25137 25238
CHAIN 25238 25339
CHAIN 25339 25440
CHAIN 25440 25541
CHAIN 25541 25642
CHAIN 25642 25743
CHAIN 25743 25844
CHAIN 25844 25945
CHAIN 25945 26046
CHAIN 26046 26147
CHAIN 26147 26248
CHAIN 26248 26349
CHAIN 26349 26450
CHAIN 26450 26551
CHAIN 26551 26652
CHAIN 26652 26753
CHAIN 26753 26854
CHAIN 26854 26955
CHAIN 26955 27056
CHAIN 27056 27157
CHAIN 27157 27258
CHAIN 27258 27359
CHAIN 27359 27460
CHAIN 27460 27561
CHAIN 27561 27662
CHAIN 27662 27763
CHAIN 27763 27864
CHAIN 27864 27965
CHAIN 27965 28066
CHAIN 28066 28167
CHAIN 28167 28268
CHAIN 28268 28369
CHAIN 28369 28470
CHAIN 28470 28571
CHAIN 28571 28672
CHAIN 28672 28773
CHAIN 28773 28874
CHAIN 28874 28975
CHAIN 28975 29076
CHAIN 29076 29177
CHAIN 29177 29278
CHAIN 29278 29379
CHAIN 29379 29480
CHAIN 29480 29581
CHAIN 29581 29682
CHAIN 29682 29783
CHAIN 29783 29884
CHAIN 29884 29985
CHAIN 29985 30086
CHAIN 30086 30187
CHAIN 30187 30288
CHAIN 30288 30389
CHAIN 30389 30490
CHAIN 30490 30591
CHAIN 30591 30692
CHAIN 30692 30793
CHAIN 30793 30894
CHAIN 30894 30995
CHAIN 30995 31096
CHAIN 31096 31197
CHAIN 31197 31298
CHAIN 31298 31399
CHAIN 31399 31400
CHAIN 31400 31501
CHAIN 31501 31602
CHAIN 31602 31703
CHAIN 31703 31804
CHAIN 31804 31905
CHAIN 31905 32006
CHAIN 32006 32107
CHAIN 32107 32208
CHAIN 32208 32309
CHAIN 32309 32410
CHAIN 32410 32511
CHAIN 32511 32612
CHAIN 32612 32713
CHAIN 32713 32814
CHAIN 32814 32915
CHAIN 32915 33016
CHAIN 33016 33117
CHAIN 33117 33218
CHAIN 33218 33319
CHAIN 33319 33420
CHAIN 33420 33521
CHAIN 33521 33622
CHAIN 33622 33723
CHAIN 33723 33824
CHAIN 33824 33925
CHAIN 33925 34026
CHAIN 34026 34127
CHAIN 34127 34228
CHAIN 34228 34329
CHAIN 34329 34430
CHAIN 34430 34531
CHAIN 34531 34632
CHAIN 34632 34733
CHAIN 34733 34834
CHAIN 34834 34935
CHAIN 34935 35036
CHAIN 35036 35137
CHAIN 35137 35238
CHAIN 35238 35339
CHAIN 35339 35440
CHAIN 35440 35541
CHAIN 35541 35642
CHAIN 35642 35743
CHAIN 35743 35844
CHAIN 35844 35945
CHAIN 35945 36046
CHAIN 36046 36147
CHAIN 36147 36248
CHAIN 36248 36349
CHAIN 36349 36450
CHAIN 36450 36551
CHAIN 36551 36652
CHAIN 36652 36753
CHAIN 36753 36854
CHAIN 36854 36955
CHAIN 36955 37056
CHAIN 37056 37157
CHAIN 37157 37258
CHAIN 37258 37359
CHAIN 37359 37460
CHAIN 37460 37561
CHAIN 37561 37662
CHAIN 37662 37763
CHAIN 37763 37864
CHAIN 37864 37965
CHAIN 37965 38066
CHAIN 38066 38167
CHAIN 38167 38268
CHAIN 38268 38369
CHAIN 38369 38470
CHAIN 38470 38571
CHAIN 38571 38672
CHAIN 38672 38773
CHAIN 38773 38874
CHAIN 38874 38975
CHAIN 38975 39076
CHAIN 39076 39177
CHAIN 39177 39278
CHAIN 39278 39379
CHAIN 39379 39480
CHAIN 39480 39581
CHAIN 39581 39682
CHAIN 39682 39783
CHAIN 39783 39884
CHAIN 39884 39985
CHAIN 39985 40086
CHAIN 40086 40187
CHAIN 40187 40288
CHAIN 40288 40389
CHAIN 40389 40490
CHAIN 40490 40591
CHAIN 40591 40692
CHAIN 40692 40793
CHAIN 40793 40894
CHAIN 40894 40995
CHAIN 40995 41096
CHAIN 41096 41197
CHAIN 41197 41298
CHAIN 41298 41399
CHAIN 41399 41400
CHAIN 41400 41501
CHAIN 41501 41602
CHAIN 41602 41703
CHAIN 41703 41804
CHAIN 41804 41905
CHAIN 41905 42006
CHAIN 42006 42107
CHAIN 42107 42208
CHAIN 42208 42309
CHAIN 42309 42410
CHAIN 42410 42511
CHAIN 42511 42612
CHAIN 42612 42713
CHAIN 42713 42814
CHAIN 42814 42915
CHAIN 42915 43016
CHAIN 43016 43117
CHAIN 43117 43218
CHAIN 43218 43319
CHAIN 43319 43420
CHAIN 43420 43521
CHAIN 43521 43622
CHAIN 43622 43723
CHAIN 43723 43824
CHAIN 43824 43925
CHAIN 43925 44026
CHAIN 44026 44127
CHAIN 44127 44228
CHAIN 44228 44329
CHAIN 44329 44430
CHAIN 44430 44531
CHAIN 44531 44632
CHAIN 44632 44733
CHAIN 44733 44834
CHAIN 44834 44935
CHAIN 44935 45036
CHAIN 45036 45137
CHAIN 45137 45238
CHAIN 45238 45339
CHAIN 45339 45440
CHAIN 45440 45541
CHAIN 45541 45642
CHAIN 45642 45743
CHAIN 45743 45844
CHAIN 45844 45945
CHAIN 45945 46046
CHAIN 46046 46147
CHAIN 46147 46248
CHAIN 46248 46349
CHAIN 46349 46450
CHAIN 46450 46551
CHAIN 46551 46652
CHAIN 46652 46753
CHAIN 46753 46854
CHAIN 46854 46955
CHAIN 46955 47056
CHAIN 47056 47157
CHAIN 47157 47258
CHAIN 47258 47359
CHAIN 47359 47460
CHAIN 47460 47561
CHAIN 47561 47662
CHAIN 47662 47763
CHAIN 47763 47864
CHAIN 47864 47965
CHAIN 47965 48066
CHAIN 48066 48167
CHAIN 48167 48268
CHAIN 48268 48369
CHAIN 48369 48470
CHAIN 48470 48571
CHAIN 48571 48672
CHAIN 48672 48773
CHAIN 48773 48874
CHAIN 48874 48975
CHAIN 48975 49076
CHAIN 49076 49177
CHAIN 49177 49278
CHAIN 49278 49379
CHAIN 49379 49480
CHAIN 49480 49581
CHAIN 49581 49682
CHAIN 49682 49783
CHAIN 49783 49884
CHAIN 49884 49985
CHAIN 49985 50086
CHAIN 50086 50187
CHAIN 50187 50288
CHAIN 50288 50389
CHAIN 50389 50490
CHAIN 50490 50591
CHAIN 50591 50692
CHAIN 50692 50793
CHAIN 50793 50894
CHAIN 50894 50995
CHAIN 50995 51096
CHAIN 51096 51197
CHAIN 51197 51298
CHAIN 51298 51399
CHAIN 51399 51400
CHAIN 51400 51501
CHAIN 51501 51602
CHAIN 51602 51703
CHAIN 51703 51804
CHAIN 51804 51905
CHAIN 51905 52006
CHAIN 52006 52107
CHAIN 52107 52208
CHAIN 52208 52309
CHAIN 52309 52410
CHAIN 52410 52511
CHAIN 52511 52612
CHAIN 52612 52713
CHAIN 52713 52814
CHAIN 52814 52915
CHAIN 52915 53016
CHAIN 53016 53117
CHAIN 53117 53218
CHAIN 53218 53319
CHAIN 53319 53420
CHAIN 53420 53521
CHAIN 53521 53622
CHAIN 53622 53723
CHAIN 53723 53824
CHAIN 53824 53925
CHAIN 53925 54026
CHAIN 54026 54127
CHAIN 54127 54228
CHAIN 54228 54329
CHAIN 54329 54430
CHAIN 54430 54531
CHAIN 54531 54632
CHAIN 54632 54733
CHAIN 54733 54834
CHAIN 54834 54935
CHAIN 54935 55036
CHAIN 55036 55137
CHAIN 55137 55238
CHAIN 55238 55339
CHAIN 55339 55440
CHAIN 55440 55541
CHAIN 55541 55642
CHAIN 55642 55743
CHAIN 55743 55844
CHAIN 55844 55945
CHAIN 55945 56046
CHAIN 56046 56147
CHAIN 56147 56248
CHAIN 56248 56349
CHAIN 56349 56450
CHAIN 56450 56551
CHAIN 56551 56652
CHAIN 56652 56753
CHAIN 56753 56854
CHAIN 56854 56955
CHAIN 56955 57056
CHAIN 57056 57157
CHAIN 57157 57258
CHAIN 57258 57359
CHAIN 57359 57460
CHAIN 57460 57561
CHAIN 57561 57662
CHAIN 57662 57763
CHAIN 57763 57864
CHAIN 57864 57965
CHAIN 57965 58066
CHAIN 58066 58167
CHAIN 58167 58268
CHAIN 58268 58369
CHAIN 58369 58470
CHAIN 58470 58571
CHAIN 58571 58672
CHAIN 58672 58773
CHAIN 58773 58874
CHAIN 58874 58975
CHAIN 58975 59076
CHAIN 59076 59177
CHAIN 59177 59278
CHAIN 59278 59379
CHAIN 59379 59480
CHAIN 59480 59581
CHAIN 59581 59682
CHAIN 59682 59783
CHAIN 59783 59884
CHAIN 59884 59985
CHAIN 59985 60086
CHAIN 60086 60187
CHAIN 60187 60288
CHAIN 60288 60389
CHAIN 60389 60490
CHAIN 60490 60591
CHAIN 60591 60692
CHAIN 60692 60793
CHAIN 60793 60894
CHAIN 60894 60995
CHAIN 60995 61096
CHAIN 61096 61197
CHAIN 61197 61298
CHAIN 61298 61399
CHAIN 61399 61400
CHAIN 61400 61501
CHAIN 61501 61602
CHAIN 61602 61703
CHAIN 61703 61804
CHAIN 61804 61905
CHAIN 61905 62006
CHAIN 62006 62107
CHAIN 62107 62208
CHAIN 62208 62309
CHAIN 62309 62410
CHAIN 62410 62511
CHAIN 62511 62612
CHAIN 62612 62713
CHAIN 62713 62814
CHAIN 62814 62915
CHAIN 62915 63016
CHAIN 63016 63117
CHAIN 63117 63218
CHAIN 63218 63319
CHAIN 63319 63420
CHAIN 63420 63521
CHAIN 63521 63622
CHAIN 63622 63723
CHAIN 63723 63824
CHAIN 63824 63925
CHAIN 63925 64026
CHAIN 64026 64127
CHAIN 64127 64228
CHAIN 64228 64329
CHAIN 64329 64430
CHAIN 64430 64531
CHAIN 64531 64632
CHAIN 64632 64733
CHAIN 64733 64834
CHAIN 64834 64935
CHAIN 64935 65036
CHAIN 65036 65137
CHAIN 65137 65238
CHAIN 65238 65339
CHAIN 65339 65440
CHAIN 65440 65541
CHAIN 65541 65642
CHAIN 65642 65743
CHAIN 65743 65844
CHAIN 65844 65945
CHAIN 65945 66046
CHAIN 66046 66147
CHAIN 66147 66248
CHAIN 66248 66349
CHAIN 66349 66450
CHAIN 66450 66551
CHAIN 66551 66652
CHAIN 66652 66753
CHAIN 66753 66854
CHAIN 66854 66955
CHAIN 66955 67056
CHAIN 67056 67157
CHAIN 67157 67258
CHAIN 67258 67359
CHAIN 67359 67460
CHAIN 67460 67561
CHAIN 67561 67662
CHAIN 67662 67763
CHAIN 67763 67864
CHAIN 67864 67965
CHAIN 67965 68066
CHAIN 68066 68167
CHAIN 68167 68268
CHAIN 68268 68369
CHAIN 68369 68470
CHAIN 68470 68571
CHAIN 68571 68672
CHAIN 68672 68773
CHAIN 68773 68874
CHAIN 68874 68975
CHAIN 68975 69076
CHAIN 69076 69177
CHAIN 69177 69278
CHAIN 69278 69379
CHAIN 69379 69480
CHAIN 69480 69581
CHAIN 69581 69682
CHAIN 69682 69783
CHAIN 69783 69884
CHAIN 69884 69985
CHAIN 69985 70086
CHAIN 70086 70187
CHAIN 70187 70288
CHAIN 70288 70389
CHAIN 70389 70490
CHAIN 70490 70591
CHAIN 70591 70692
CHAIN 70692 70793
CHAIN 70793 70894
CHAIN 70894 70995
CHAIN 70995 71096
CHAIN 71096 71197
CHAIN 71197 71298
CHAIN 71298 71399
CHAIN 71399 71400
CHAIN 71400 71501
CHAIN 71501 71602
CHAIN 71602 71703
CHAIN 71703 71804
CHAIN 71804 71905
CHAIN 71905 72006
CHAIN 72006 72107
CHAIN 72107 72208
CHAIN 72208 72309
CHAIN 72309 72410
CHAIN 72410 72511
CHAIN 72511 72612
CHAIN 72612 72713
CHAIN 72713 72814
CHAIN 72814 72915
CHAIN 72915 73016
CHAIN 73016 73117
CHAIN 73117 73218
CHAIN 73218 73319
CHAIN 73319 73420
CHAIN 73420 73521
CHAIN 73521 73622
CHAIN 73622 73723
CHAIN 73723 73824
CHAIN 73824 73925
CHAIN 73925 74026
CHAIN 74026 74127
CHAIN 74127 74228
CHAIN 74228 74329
CHAIN 74329 74430
CHAIN 74430 74531
CHAIN 74531 74632
CHAIN 74632 74733
CHAIN 74733 74834
CHAIN 74834 74935
CHAIN 74935 75036
CHAIN 75036 75137
CHAIN 75


```

FT STRAND 560 562
FT STRAND 567 573
FT TURN 575 576
FT STRAND 577 577
FT TURN 582 583
FT STRAND 586 586

Query Match 5.7%; Score 88; DB 1; Length 3414;
Best Local Similarity 23.0%; Pred. No. 42;
Matches 59; Conservative 28; Mismatches 86; Indels 84; Gaps 12;

QY 38 HLTPLQD-----WAAAGFLDAVAVEPIFSGMEVXITTWGADTACGDIISGLFVSA 90
DB 1442 HLTLEKBERMMATWLIAGL-----AASAIHWSGILVWGLMTLTEMASRRSDLVFSG 1496
QY 91 RRGREILLGPAADNEGQGMRLAPITAYSOQTRGLCITSLGRDKNGVEGVVST 150
DB 1497 QGGEREDRPEVVDGV-YRIFSGLFWGQ-----NOVG-----VGY 1532
QY 151 ATGSPFLATCVNGVCMVTFHGAG--SKTLGPKGPITQMYTNVDQV-----GW 1597
DB 1533 GSKGVLT-----WMHTYRGAAISIDDAVAGP-----YVADVEDVVCYGAMSLEKV 1581
QY 198 QA-----PRGARSMTPCTCGSSDLYVTRADVPVRRRGDSRLSPRPVSLKG 249
DB 1582 KGETVQVHAPFPG-RAHEVHOCQPGELIDT-----GRKLGAIPIDLKVG 1625
QY 250 SSGGPLLCPSGHVGIF 266
DB 1626 TSGSPILMAQGVVGLV 1642

RESULT 15
POLG TBESVS STANDARD; PRT; 3412 AA.
AC P07720; P07721; Q88475; Q88476; Q88477; Q88478; Q88479; Q88877;
AC Q88878; Q88879;
AC 01-APR-1988 (Rel. 07, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Genome polypotein [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase
DE (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)
DE (NS5)].
OS Tick-borne encephalitis virus (strain Sofjin) (TBEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_Taxid=11087;
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE=90101381; PubMed=2136778;
RA Pletnev A.G., Yamschikov V.F., Bilinov V.M.;
RT "Nucleotide sequence of the genome and complete amino acid sequence
RT of the polypotein of tick-borne encephalitis virus.";
RL Virology 174:250-263 (1990).
RN 12
RP SEQUENCE OF 1-1190 FROM N.A.
RA MEDLINE=88313988; PubMed=2970626;
RA Yamschikov V.F., Pletnev A.G.;
RT "Nucleotide sequence of the genome region encoding the structural
RT proteins and the NS1 protein of the tick borne encephalitis virus.";
RL Nucleic Acids Res. 16:7750-7750 (1988).
RN 13
RP SEQUENCE OF 1-683 AND 777-1002 FROM N.A.
RA MEDLINE=86220766; PubMed=3709796;
RA Pletnev A.G., Yamschikov V.F., Bilinov V.M.;
RT "Tick-borne encephalitis virus genome. The nucleotide sequence coding
RT for virion structural proteins.";
RL FEBS Lett. 200:317-321 (1986).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function.
CC NS3 and NS5 may play a role in the viral RNA replication.

```

```

CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypotein, commonly with Asp or Glu in the P6
CC position; Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
CC (RNA) (N).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and RNA.
CC -1- MISCELLANEOUS: The nonstructural protein NS1 presents two
CC alternative cleavage sites for its C-terminus, which may define a
CC soluble or a membrane-bound form of NS1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DB EMBL; X07755; CAA30581.1; -
DB EMBL; X03870; CAA27501.1; -
DB EMBL; X03870; CAA27501.1; ALT_SEQ.
DB EMBL; X03870; CAA27502.1; ALT_SEQ.
DB EMBL; X03870; CAA27503.1; ALT_SEQ.
DB EMBL; X03871; CAA27505.1; -
DB MEROPS; S07.UPW; -.
DR InterPro; IPR009003; Cys_ser_trypsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001122; Flavi_capsidC.
DR InterPro; IPR000336; Flavi_glycoprote.
DR InterPro; IPR000069; Flavi_M.
DR InterPro; IPR001157; Flavi_NS1.
DR InterPro; IPR000752; Flavi_NS2A.
DR InterPro; IPR000487; Flavi_NS2B.
DR InterPro; IPR000404; Flavi_NS4A.
DR InterPro; IPR001528; Flavi_NS4B.
DR InterPro; IPR000208; Flavi_NS5.
DR InterPro; IPR002535; Flavi_propep.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR001650; Peptidase_S7.
DR InterPro; IPR007085; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_DS_Vir.
DR InterPro; IPR002877; Rnm_FtsJ.
DR Pfam; PF01003; Flavi_capsid; 1.
DR Pfam; PF02832; Flavi_glycoprot; 1.
DR Pfam; PF00869; Flavi_glycoprot; 1.
DR Pfam; PF00949; Flavi_helicase; 1.
DR Pfam; PF01004; Flavi_M; 1.
DR Pfam; PF00948; Flavi_NS1; 1.
DR Pfam; PF01005; Flavi_NS2A; 1.
DR Pfam; PF01002; Flavi_NS2B; 1.
DR Pfam; PF01350; Flavi_NS4A; 1.
DR Pfam; PF01349; Flavi_NS4B; 1.
DR Pfam; PF00972; Flavi_NS5; 1.
DR Pfam; PF01570; Flavi_propep; 1.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF00271; helicase_C; 1.
DR ProDom; PD001556; Flavi_glycoprote; 1.
DR ProDom; PD001496; Flavi_NS1; 1.
DR PROSITE; PS00660; DEAH_ATP_Helicase; FALSE NEG.
KM Core protein; Coat protein; Envelope protein; Hydrophobic; Helicase;
KM ATP-binding; Transmembrane; Nonstructural protein.
FT CHAIN 1 112
FT PROPEP 113 205
FT CHAIN 206 280
FT CHAIN 281 776
FT CHAIN 777 21128
ENVELOPE GLYCOPROTEIN M.
MAJOR ENVELOPE PROTEIN E.
NONSTRUCTURAL PROTEIN NS1 (OR 1190).

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 27.3102 Seconds

(without alignments)
3373.509 Million cell updates/sec

Title: US-10-650-585-15
Perfect score: 1532
Sequence: 1 AAGGIRACMVLVRKAGSHY.....RGVAKAVDFIPVSMETWR 292

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 segs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1499	97.8	3010	12 Q9J3H7	Q9J3H7 hepatitis c
2	1499	97.8	3010	12 Q9D7D6	Q9D7D6 hepatitis c
3	1494	97.5	3010	12 P88803	P88803 hepatitis c
4	1494	97.5	3010	12 Q68826	Q68826 hepatitis c
5	1490	97.3	3010	12 Q9D7E6	Q9D7E6 hepatitis c
6	1490	97.3	3010	12 P90191	P90191 hepatitis c
7	1490	97.3	3013	12 Q9J3H4	Q9J3H4 hepatitis c
8	1489	97.2	3008	12 Q9J3F4	Q9J3F4 hepatitis c
9	1489	97.2	3010	12 Q9Q1Y3	Q9Q1Y3 hepatitis c
10	1486	96.9	3010	12 Q9J3F9	Q9J3F9 hepatitis c
11	1485	96.9	3010	12 Q9J3H5	Q9J3H5 hepatitis c
12	1485	96.9	3010	12 Q9D7E4	Q9D7E4 hepatitis c
13	1485	96.9	3010	12 Q807P3	Q807P3 hepatitis c
14	1482	96.7	3010	12 Q9J3I1	Q9J3I1 hepatitis c
15	1481	96.7	361	12 Q70822	Q70822 hepatitis c
16	1481	96.7	3010	12 Q9D7D7	Q9D7D7 hepatitis c

17	1481	96.7	3010	12 Q81541	Q81541 hepatitis c
18	1481	96.7	3011	12 Q9D7E2	Q9D7E2 hepatitis c
19	1480	96.6	361	12 Q70815	Q70815 hepatitis c
20	1480	96.6	3010	12 Q9J3G3	Q9J3G3 hepatitis c
21	1480	96.6	3010	12 Q9Q1Y3	Q9Q1Y3 hepatitis c
22	1480	96.6	3010	12 Q9J3H3	Q9J3H3 hepatitis c
23	1480	96.6	3010	12 Q9Q1Y4	Q9Q1Y4 hepatitis c
24	1480	96.6	3014	12 Q9D7E0	Q9D7E0 hepatitis c
25	1479	96.5	3010	12 Q9Q1Y7	Q9Q1Y7 hepatitis c
26	1479	96.5	3010	12 Q9J3I0	Q9J3I0 hepatitis c
27	1479	96.5	3010	12 Q9Q1Y8	Q9Q1Y8 hepatitis c
28	1479	96.5	3010	12 Q9J3H6	Q9J3H6 hepatitis c
29	1478	96.5	3010	12 Q9J3G6	Q9J3G6 hepatitis c
30	1478	96.5	3010	12 Q9Q1X6	Q9Q1X6 hepatitis c
31	1477	96.4	186	12 Q81755	Q81755 hepatitis c
32	1477	96.4	284	12 Q81817	Q81817 hepatitis c
33	1477	96.4	3010	12 Q9Q1X4	Q9Q1X4 hepatitis c
34	1477	96.4	3010	12 P89966	P89966 hepatitis c
35	1477	96.4	3010	12 Q9Q1X6	Q9Q1X6 hepatitis c
36	1477	96.4	3010	12 Q9Q1X5	Q9Q1X5 hepatitis c
37	1476	96.3	3010	12 Q9J3H9	Q9J3H9 hepatitis c
38	1476	96.3	3010	12 Q9J3H2	Q9J3H2 hepatitis c
39	1476	96.3	3010	12 Q9D7F0	Q9D7F0 hepatitis c
40	1476	96.3	3010	12 Q9Q1Y5	Q9Q1Y5 hepatitis c
41	1475	96.3	3010	12 Q9D7E9	Q9D7E9 hepatitis c
42	1475	96.3	3010	12 Q68788	Q68788 hepatitis c
43	1474	96.2	3010	12 Q99A02	Q99A02 hepatitis c
44	1473	96.1	3010	12 Q9W0X2	Q9W0X2 hepatitis c
45	1473	96.1	3010	12 Q9W0X2	Q9W0X2 hepatitis c

ALIGNMENTS

RESULT 1
ID Q9J3H7 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepatitis virus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD15.
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease progression."
RT Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
CC
CC EMBL; AF207756; AAF65946.1; -.
DR PIR; A61196; A61196.
DR PIR; PQ0246; PQ0246.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0329; PS0329.
DR HSSP; P26663; IUXP.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO:0016740; F:transferase activity; IEA.
 DR GO:0006118; P:electron transport; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO:0006350; P:transcription; IEA.
 DR GO:0019079; P:viral genome replication; IEA.
 DR GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_transf. IEA.
 DR InterPro: IPR00345; CysC_heme_BS.
 DR InterPro: IPR01410; DEAD.
 DR InterPro: IPR00522; HCV_capsid.
 DR InterPro: IPR00521; HCV_core.
 DR InterPro: IPR00519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR00745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_NS5a.
 DR InterPro: IPR004108; peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVlr.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART; SMO0487; DEXdc; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327365 MW; D86537317FFA106 CRC64;

Query Match 97.8%; Score 1499; DB 12; Length 3010;
 Best Local Similarity 96.2%; Pred. No. 1.3e-124;
 Matches 281; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

DR 1 AAGGIRACMTVRKAGGHHYQMAFMKLAALGTIVYDHLTPLODMARGLRLAVAVEPY 60
 DB 915 AAGGIRACMTVRKAGGHHYQMAFMKLAALGTIVYDHLTPLODMARGLRLAVAVEPY 974
 QY 61 IFSDMEVYKITTGADTACGDIISGLFVSARGRHEILGADNFEQGRLLAPITAYSQ 120
 DB 975 VFSDMEVYKITTGADTACGDIISGLFVSARGRHEILGADNFEQGRLLAPITAYSQ 1034
 QY 121 QTRGLGCIITSLGRKNGVEGVVSTATOSFIATCVNGVCMVTFHAGSXTLAGPK 180
 DB 1035 QTRGLGCIITSLGRKNGVEGVVSTATOSFIATCVNGVCMVTFHAGSXTLAGPK 1094
 QY 181 GPITQMTYVQDILVGNQAPPGARSMPTCTGSSDLYLVTRHADVIIPRRRGDSRGLLS 240
 DB 1095 GPITQMTYVQDILVGNQAPPGARSMPTCTGSSDLYLVTRHADVIIPRRRGDSRGLLS 1154
 QY 241 PRPVSYKSGSGGGLCPGSHAVIFRAAVCTRGVAFAVVFIVESMETMR 292
 DB 1155 PRPVSYKSGSGGGLCPGSHAVIFRAAVCTRGVAFAVVFIVESMETMR 1206

RESULT 2
 Q9DPTD6 PRELIMINARY; PRT; 3010 AA.
 AC Q9DPTD6
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 OS Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV221;
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
 RA Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijioka M.,
 RA Mishiro S.;
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
 with hepatocellular carcinoma: the 'progression score' revisited.",
 RL Submitted (Sep-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; AB049101; BAB18914.1; -.
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.
 DR PIR; P80329; P80329.
 DR HSSD; P26663; 1UXP.
 DR GO:0016021; C:integral to membrane; IEA.
 DR GO:0019028; C:viral capsid; IEA.
 DR GO:0019031; C:viral envelope; IEA.
 DR GO:0005524; F:ATP binding; IEA.
 DR GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO:0005489; F:electron transporter activity; IEA.
 DR GO:0016787; F:hydrolyase activity; IEA.
 DR GO:0003723; F:RNA binding; IEA.
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO:0005198; F:structural molecule activity; IEA.
 DR GO:0016740; F:transferase activity; IEA.
 DR GO:0006118; F:electron transport; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO:0006350; P:transcription; IEA.
 DR GO:0019079; P:viral genome replication; IEA.
 DR GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_transf.
 DR InterPro: IPR00345; CysC_heme_BS.
 DR InterPro: IPR01410; DEAD.
 DR InterPro: IPR00522; HCV_capsid.
 DR InterPro: IPR00521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR00745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_NS5a.
 DR InterPro: IPR004108; peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVlr.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART; SMO0487; DEXdc; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolyase; Nonstructural protein; Polypeptide; Transmembrane.
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327108 MW; DE182D810E7BEE4 CRC64;

Query Match 97.8%; Score 1499; DB 12; Length 3010;

Best Local Similarity 96.6%; Pred. No. 1,36-124;
Matches 282; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 60
Db 915 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 974
QY 61 IFSMEVKTITWGADTAACGDIISGLPVASRGREIILGPADNFEQGWRLAIPITAYSQ 120
Db 975 VFSMEVKTITWGADTAACGDIISGLPVASRGREIILGPADNFEQGWRLAIPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAAGPK 180
Db 1035 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAAGPK 1094
QY 181 GPITQMTNVDDLVGMQAPPGASMTPTCGSSDLYLVTRADVIIVRRRDSGSLLS 240
Db 1095 GPITQMTNVDDLVGMQAPPGASMTPTCGSSDLYLVTRADVIIVRRRDSGSLLS 1154
QY 241 PRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 292
Db 1155 PRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 1206

RESULT 3

P88803 PRELIMINARY; PRT; 3010 AA.

AC P88803;
DT 01-MAY-1997 (TREMELREL. 03, Created)
DT 01-MAY-1997 (TREMELREL. 03, Last sequence update)
DT 01-OCT-2003 (TREMELREL. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-1b;
RA Enomoto N.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-1b;
RA MEDLINE=95340824; PubMed=7542279;
RA Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T., Yamamoto C., Izumi N., Marumo F., Sato C.;
RA "Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b";
RT J. Clin. Invest. 96:224-230 (1995).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL: D50484; BAA09075.1; -
DR PTR: A61196; A61196.
DR HSBP; P28663; INS3.
DR GO: GO:0016021; C:Integral to membrane; IEA.
DR GO: GO:0019028; C:Viral capsid; IEA.
DR GO: GO:0019031; C:Viral envelope; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003368; F:RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006350; P:proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR GO: GO:0019079; P:viral genome replication; IEA.
DR GO: GO:0019087; P:viral transformation; IEA.
DR InterPro: IPR009003; Cys_ser_lypsin.
DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5b.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004109; peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSTr.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1
FT CHAIN 191
FT CHAIN 192 383 E1.
FT CHAIN 384 809 E2.
FT CHAIN 810 1026 NS2.
FT CHAIN 1027 1657 NS3.
FT CHAIN 1658 1711 NS4a.
FT CHAIN 1712 1972 NS4b.
FT CHAIN 1973 2419 NS5a.
FT CHAIN 2420 3010 NS5b.
SQ SEQUENCE 3010 AA; 327332 MW; 5F81505783FEFFB8 CRC64;

Query Match 97.5%; Score 1494; DB 12; Length 3010;
Best Local Similarity 96.2%; Pred. No. 3,66-124;
Matches 281; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 60
Db 915 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 974
QY 61 IFSMEVKTITWGADTAACGDIISGLPVASRGREIILGPADNFEQGWRLAIPITAYSQ 120
Db 975 VFSMEVKTITWGADTAACGDIISGLPVASRGREIILGPADNFEQGWRLAIPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAAGPK 180
Db 1035 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAAGPK 1094
QY 181 GPITQMTNVDDLVGMQAPPGASMTPTCGSSDLYLVTRADVIIVRRRDSGSLLS 240
Db 1095 GPITQMTNVDDLVGMQAPPGASMTPTCGSSDLYLVTRADVIIVRRRDSGSLLS 1154
QY 241 PRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 292
Db 1155 PRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 1206

RESULT 4

O68826 PRELIMINARY; PRT; 3010 AA.

AC O68826;
DT 01-NOV-1996 (TREMELREL. 01, Created)
DT 01-NOV-1996 (TREMELREL. 01, Last sequence update)
DT 01-OCT-2003 (TREMELREL. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.

DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4A.
 DR InterPro: IPR001490; HCV NS4B.
 DR InterPro: IPR002868; HCV NS5A.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA pol DS PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00480; HELIC; 1.
 DR PROSITE: PS00190; CYTOCHROME C; 1.
 DR ATP-binding; Coat protein; Glycoprotein; Helicase;
 KM Hydroxylase; Nonstructural protein; Polyprotein;
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;

Query Match 97.3%; Score 1490; DB 12; Length 3010;
 Best Local Similarity 95.9%; Pred. No. 8.1e-124;
 Matches 280; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
 QY 1 AAGLRACMTVRKAGGHHVYOMAFMKALATGTGVYDHLPLDPMAHAGRDAAVEPV 60
 DB 915 AAGLRACMTVRKAGGHHVYOMAFMKALATGTGVYDHLPLDPMAHAGRDAAVEPV 974
 QY 61 IFSMEVKKITWGADTAACGDIISGLFVSARGRHEILLGPADNFEQGMWLLAPITAYSQ 120
 DB 975 VFSMEVKKITWGADTAACGDIISGLFVSARGRHEILLGPADNFEQGMWLLAPITAYSQ 1034
 QY 121 QTRSLCCITSLTGRKNVGEVGVYVATQSFATCNQGVWTFHAGSKTLAPK 180
 DB 1035 QTRSLCCITSLTGRKNVGEVGVYVATQSFATCNQGVWTFHAGSKTLAPK 1094
 QY 181 GPITQWYTNVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVLPVRRRGRSGLLS 240
 DB 1095 GPITQWYTNVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVLPVRRRGRSGLLS 1154
 QY 241 PRPVSYLKSGSGGPLLCPGSHAVIFRAVCTRGVAAVDFPIPVESMETTMR 292
 DB 1155 PRPVSYLKSGSGGPLLCPGSHAVIFRAVCTRGVAAVDFPIPVESMETTMR 1206

RESULT 6

ID P90191 PRELIMINARY; PRT; 3010 AA.
 AC P90191;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV-1b;
 RA Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=HCV-1b;
 RX MEDLINE=95340824; PubMed=7542279;
 RA Etomoto N., Sakuma T., Asahina Y., Kuroski M., Murakami T.,
 RA Yamamoto C., Izumi N., Maruno F., Sato C.;
 RT "Comparison of full-length sequences of interferon-sensitive and
 RT resistant hepatitis C virus 1b."
 RL J. Clin. Invest. 96:224-230(1995).
 CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; D50482; BAA09073.1; -.
 DR PIR; A61196; A61196.
 DR PIR; P00254; P00254.
 DR PIR; P00804; P00804.
 DR PIR; P80329; P80329.
 DR PDB; 1DXW; 12-JAN-01.
 DR GO; GO:0016021; C:Integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:003723; F:RNA binding; IEA.
 DR GO; GO:0003668; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; F:transcription; IEA.
 DR GO; GO:0019079; F:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_trypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA pol DS PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 FT CHAIN 1 191
 FT CHAIN 192 383
 FT CHAIN 384 809
 FT CHAIN 810 1026
 FT CHAIN 1027 1657
 FT CHAIN 1658 1711
 FT CHAIN 1712 1972
 FT CHAIN 1973 2419
 FT CHAIN 2420 3010
 SQ SEQUENCE 3010 AA; 327438 MW; 5F15AC675A0C8268 CRC64;

Query Match

97.3%; Score 1490; DB 12; Length 3010;

Best Local Similarity 96.2%; Pred. No. 8,1e-124; Matches 281; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRAAGGHVYQMAFMKLAALTGTYYDHLTLPLOMAHAGLRDLAAVEPV 60

DB 915 AAGLIRACMLVRAAGGHVYQMAFMKLAALTGTYYDHLTLPLOMAHAGLRDLAAVEPV 974

QY 61 IFSDMEVKIITWGDADTAACGDIISGLPVSARGREIILGPADNFGCGMRLAIPITAYSQ 120

DB 975 VFSDMEKIIITWGDADTAACGDIISGLPVSARGREIILGPADNFGCGMRLAIPITAYSQ 1034

QY 121 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTHGAGSKTLAEPK 180

DB 1035 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTHGAGSKTLAEPK 1094

QY 181 GPITOMTNNVDOLVGMQAPPGARSMPCTCGSSDLYLVTRHADVIPIVRRGDSRGLLS 240

DB 1095 GPITOMTNNVDOLVGMQAPPGARSMPCTCGSSDLYLVTRHADVIPIVRRGDSRGLLS 1154

QY 241 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAVDFIPVESMETTMR 292

DB 1155 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAVDFIPVESMETTMR 1206

RESULT 7

Q9J3H4 PRELIMINARY; PRT; 3013 AA.

AC Q9J3H4; 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Genome polypotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

CC NCBI_Taxid=11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MD18;

RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.; "Characteristics of hepatitis C viral genome associated with disease progression."

RT submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).

CC EMBL; AF207759; AAR65949.1; -

DB PIR; A61196; A61196.

DR PIR; P00804; P00804.

DR PIR; PS0329; PS0329.

DR HSPF; P26663; TXP.

DR GO; GO:0016021; C: integral to membrane; IEA.

DR GO; GO:0015028; C: viral capsid; IEA.

DR GO; GO:0019031; C: viral envelope; IEA.

DR GO; GO:0005824; F: ATP binding; IEA.

DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F: electron transporter activity; IEA.

DR GO; GO:0003723; F: RNA binding; IEA.

DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008265; F: serine-type peptidase activity; IEA.

DR GO; GO:0005198; F: structural molecule activity; IEA.

DR GO; GO:0016740; F: transferase activity; IEA.

DR GO; GO:0006108; P: proteolysis and peptidolysis; IEA.

DR GO; GO:0006508; P: transcription; IEA.

DR GO; GO:0019079; P: viral genome replication; IEA.

DR GO; GO:0019087; P: viral genome transformation; IEA.

DR InterPro; IPR003003; Cys Ser LysPsn.

DR InterPro; IPR000345; Cys_heme_BS.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV NS5b.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR004109; peptidase C29.

DR InterPro; IPR007095; RNA pol DS_PS.

DR InterPro; IPR007094; RNA pol_PSV1r.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; Helicase_C; 1.

DR Pfam; PF00998; Viral_RDRP; 1.

DR ProDom; PD186062; HCV NS1; 1.

DR SMART; SMO0487; DEXDC; 1.

DR PROSITE; PS00190; CYTOCHROME C; 1.

DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; RNA-directed RNA polymerase; Transferase; Transmembrane.

KW POLYPEPTIDE; RNA-directed RNA polymerase; Transferase; Transmembrane.

SQ SEQUENCE 3013 AA; 326956 MW; E4F80FEA05C1E5 CRC64;

Query Match 97.3%; Score 1490; DB 12; Length 3013;

Best Local Similarity 96.2%; Pred. No. 8,1e-124; Matches 281; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRAAGGHVYQMAFMKLAALTGTYYDHLTLPLOMAHAGLRDLAAVEPV 60

DB 915 AAGLIRACMLVRAAGGHVYQMAFMKLAALTGTYYDHLTLPLOMAHAGLRDLAAVEPV 974

QY 61 IFSDMEVKIITWGDADTAACGDIISGLPVSARGREIILGPADNFGCGMRLAIPITAYSQ 120

DB 975 VFSDMEKIIITWGDADTAACGDIISGLPVSARGREIILGPADNFGCGMRLAIPITAYSQ 1034

QY 121 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTHGAGSKTLAEPK 180

DB 1035 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTHGAGSKTLAEPK 1094

QY 181 GPITOMTNNVDOLVGMQAPPGARSMPCTCGSSDLYLVTRHADVIPIVRRGDSRGLLS 240

DB 1095 GPITOMTNNVDOLVGMQAPPGARSMPCTCGSSDLYLVTRHADVIPIVRRGDSRGLLS 1154

QY 241 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAVDFIPVESMETTMR 292

DB 1155 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAVDFIPVESMETTMR 1206

RESULT 8

Q9J3F4 PRELIMINARY; PRT; 3008 AA.

AC Q9J3F4; 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Genome polypotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

CC NCBI_Taxid=11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MD34;

RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.; "Characteristics of hepatitis C viral genome associated with disease progression."

Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 DR EMBL; AF208024; AAE61205.1; -.
 DR PIR; A61196; A61196.
 DR PIR; PQ0246; PQ0246.
 DR PIR; PS0329; PS0329.
 DR HSSP; P26663; IJXP.
 DR GO; GO:0016021; C:Integral to membrane; IEA.
 DR GO; GO:0019028; C:Viral capsid; IEA.
 DR GO; GO:0019031; C:Viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0005489; F:electron transporter activity; IEA.
 DR GO; GO:0016787; F:hydrolyase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR InterPro; IPR009003; Cys_ser_typsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_core.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NSI.
 DR InterPro; IPR002518; HCV_NSI.
 DR InterPro; IPR000745; HCV_NS4.
 DR InterPro; IPR001490; HCV_NS4.
 DR InterPro; IPR002868; HCV_NS5.
 DR InterPro; IPR002166; HCV_NS5.
 DR InterPro; IPR001650; HCV_RDRP.
 DR InterPro; IPR004109; peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR Pfam; PF01538; HCV_NSI; 1.
 DR Pfam; PF02907; HCV_NS2; 1.
 DR Pfam; PF01006; HCV_NS4; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; viral_RDRP; 1.
 DR Pfam; PF01806; HCV_NSI; 1.
 DR SMART; SMO0487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolyase; Nonstructural protein; Polypeptide; Polypeptide;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SO SEQUENCE 3008 AA; 326834 MW; 99AE09E14C3109F4 CRC64;

Query Match 97.2%; Score 1489; DB 12; Length 3008;
 Best Local Similarity 96.2%; Pred. No. 1e-123;
 Matches 281; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGIIRACMLVRKAGHYVGMFMKLAALGTGVVHLPLOGMAGHLDLAVNVEPV 60
 DB 913 AAGIIRACMLVRKAGHYVGMFMKLAALGTGVVHLPLOGMAGHLDLAVNVEPV 972

QY 61 IFSMNEVKIITWGDATACGDIISGLPVSARGREIILGPADNPEGQWMLAPITAYSQ 120
 DB 973 VFSMNEVKIITWGDATACGDIISGLPVSARGREIILGPADNPEGQWMLAPITAYSQ 1032

QY 121 QTRGLGCIITSLTGRDKQVGEVQVSTATQSLATCVNGCVTFPHAGSKTLAGPK 180
 DB 1033 QTRGLGCIITSLTGRDKQVGEVQVSTATQSLATCVNGCVTFPHAGSKTLAGPK 1092

QY 181 GPTQWNTVDDIDVGMQAPPGARSMTPCTGSSDLVLTIRADVIIPARRRDSRGSILS 240
 DB 1093 GPTQWNTVDDIDVGMQAPPGARSLTPCTGSSDLVLTIRADVIIPARRRDSRGSILS 1152

QY 241 PRPVSYLKSSGSGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETTR 292
 DB 1153 PRPVSYLKSSGSGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETTR 1204

RESULT 9
 ID Q90IY3 PRELIMINARY; PRT; 3010 AA.
 AC Q90IY3;
 DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 CC NCBI_TaxID=11103;
 CX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=MD4-2;
 RC MEDLINE=20013325; PubMed=10544098;
 RX Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,
 RA Tazawa J., Izumi N., Maruno F., Sato C.;
 RT "Time-related changes in full-length hepatitis C virus and hepatitis
 RL virology 263:244-253(1999)."
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD4-2;
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
 RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Maruno F., Sato C.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; AF165052; AAD56187.1; -.
 DR PIR; A61196; A61196.
 DR PIR; PQ0246; PQ0246.
 DR PIR; PQ0804; PQ0804.
 DR PIR; PS0329; PS0329.
 DR HSSP; P26663; IJXP.
 DR GO; GO:0016021; C:Integral to membrane; IEA.
 DR GO; GO:0019028; C:Viral capsid; IEA.
 DR GO; GO:0019031; C:Viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR InterPro; IPR009003; Cys_ser_typsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NSI.
 DR InterPro; IPR002518; HCV_NSI.
 DR InterPro; IPR000745; HCV_NS4.

DR InterPro: IPR001490; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA pol_D5_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327087 MW; 2FE78321686D4002 CRC64;

Query Match 97.2%; Score 1489; DB 12; Length 3010;
 Best Local Similarity 95.5%; Pred. No. 1e-123;
 Matches 279; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 1 AGLIRACMLVRKAAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAEVPV 60
 DB 915 AGLIRACMLVRKAAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAEVPV 974
 QY 61 IFSDEVKIITWGADTAACGDIISGLPVSARREIILGPADNPEGQWLLAFTAYSQ 120
 DB 975 VFSMEKVIITWGADTAACGDIISGLPVSARREIILGPADNPEGQWLLAFTAYSQ 1034
 QY 121 QTRGLGCIITSLTRGRDNQVEGEVQVSTATOSFLATCVNGVMTVFHAGSKTLAGPK 180
 DB 1035 QTRGLGCIITSLTRGRDNQVEGEVQVSTATOSFLATCVNGVMTVFHAGSKTLAGPK 1094
 QY 181 GPIQWNTYNDODLVGQAPGARSMPTCTCGSSDLYLVTRHADVIVRRRGDSRGLLS 240
 DB 1095 GPIQWNTYNDODLVGQAPGARSLPTCTCGSSDLYLVTRHADVIVRRRGDSRGLLS 1154
 QY 241 PRPVSYLKSGSGGFLCPSGHAGVIFPAVCTRGVAKAVDFIVESNETTMR 292
 DB 1155 PRPVSYLKSGSGGFLCPSGHAGVIFPAVCTRGVAKAVDFIVESNETTMR 1206

RESULT 10

Q9J3F9 PRELIMINARY; PRT; 3010 AA.

AC Q9J3F9; OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=1103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ND33; Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
 RA Nagayama K., Characteristics of hepatitis C viral genome associated with disease
 RT progression.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 DR EMBL; AF207774; AAF65964.1; -.

DR PIR: A61196; A61196.
 DR PIR: P00246; P00246.
 DR PIR: P0329; P0329.
 DR HSSP: P27958; IHEI.
 DR MEROPS: S29.001; -.
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0019028; C:viral capsid; IEA.
 DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO: GO:0005489; F:electron transporter activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F:structural molecule activity; IEA.
 DR GO: GO:0016749; F:transferase activity; IEA.
 DR GO: GO:0006118; F:electron transport; IEA.
 DR GO: GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; F:transcription; IEA.
 DR GO: GO:0019079; F:viral genome replication; IEA.
 DR GO: GO:0019087; F:viral transformation; IEA.
 DR InterPro: IPR009003; Cys Ser-lysin.
 DR InterPro: IPR000345; CysC_Neme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002518; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002688; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR004109; Peptidase_C.
 DR InterPro: IPR007095; RNA pol_D5_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C_1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327102 MW; 7162C9DB93B6E0C7 CRC64;

Query Match 97.0%; Score 1486; DB 12; Length 3010;
 Best Local Similarity 95.5%; Pred. No. 1.8e-123;
 Matches 279; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 1 AGLIRACMLVRKAAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAEVPV 60
 DB 915 AGLIRACMLVRKAAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAEVPV 974
 QY 61 IFSDEVKIITWGADTAACGDIISGLPVSARREIILGPADNPEGQWLLAFTAYSQ 120
 DB 975 VFSMEKVIITWGADTAACGDIISGLPVSARREIILGPADNPEGQWLLAFTAYSQ 1034
 QY 121 QTRGLGCIITSLTRGRDNQVEGEVQVSTATOSFLATCVNGVMTVFHAGSKTLAGPK 180
 DB 1035 QTRGLGCIITSLTRGRDNQVEGEVQVSTATOSFLATCVNGVMTVFHAGSKTLAGPK 1094
 QY 181 GPIQWNTYNDODLVGQAPGARSMPTCTCGSSDLYLVTRHADVIVRRRGDSRGLLS 240

```

Db      1095 GPTQMTYNTVDLWGPAPGARSLLPCTCGSSDLVLFVHNAVIVPARRGDSRSLLS 1154
Qy      241 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 292
Db      1155 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 11
Q9J3H5 PRELIMINARY, PRT, 3010 AA.
ID Q9J3H5
AC Q9J3H5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD17;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
DR EMBL: AF207758; AAPE5948.1; -
DR PIR: A61196; A61196.
DR PIR: P00246; P00246.
DR PIR: P00254; P00254.
DR PIR: P50329; P50329.
DR HSSP: P27958; 1HE1.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0019028; C:viral capsid; IEA.
DR GO: GO:0019031; C:viral envelope; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO: GO:0005489; F:electron transporter activity; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR GO: GO:0019079; P:viral genome replication; IEA.
DR GO: GO:0019087; P:viral transformation; IEA.
DR InterPro: IPR009003; Cys_ser_lypsin.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR000745; HCV NS4a.
DR InterPro: IPR001490; HCV NS4b.
DR InterPro: IPR002868; HCV NS5a.
DR InterPro: IPR002166; HCV RdRp.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004109; Peptidase_C9.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVlr.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.

```

```

DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR POSITE: PS00190; CYTOCHROME C; 1.
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydroxylase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 326801 MW; 9FEE3D1B93B7AA4B CRC64;

Query Match 96.9%; Score 1485; DB 12; Length 3010;
Best Local Similarity 95.9%; Pred. No. 2,3e-123;
Matches 280; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy      1 AGLIPACMLVRKACGHTYVQAFMKALITGTYVDHITPLQDMAHAGLRDLAVAEVY 60
Db      915 AGLIPACMLVRKVGAGHYVQAMKMLALITGTYVDHITPLRDAVHGLRDLAVAEVY 974
Qy      61 IFSDMEVKIITMGADTAAAGDIIISGLPVASRGREIILGADNFEQGRILAPITAYSQ 120
Db      975 VESDMEVKIITMGADTAAAGDIIISGLPVASRGREIILGADNFEQGRILAPITAYSQ 1034
Qy      121 QTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVGCVTFVAGSKTLAIGK 180
Db      1035 QTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVGCVTFVAGSKTLAIGK 1094
Qy      181 GPTQMTYNTVDLWGPAPGARSLLPCTCGSSDLVLFVHNAVIVPARRGDSRSLLS 240
Db      1095 GPTQMTYNTVDLWGPAPGARSLLPCTCGSSDLVLFVHNAVIVPARRGDSRSLLS 1154
Qy      241 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 292
Db      1155 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 12
Q9DTE4 PRELIMINARY, PRT, 3010 AA.
ID Q9DTE4
AC Q9DTE4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV150;
RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA Hatahara T., Ohta Y., Kanai K., Marumo H., Baba K., Hijikata M.,
RA Yoshiro S.;
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT with hepatocellular carcinoma: the 'progression score' revisited."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
DR EMBL: AB049093; BAB1806.1; -
DR PIR: A61196; A61196.
DR PIR: P00246; P00246.
DR PIR: P00804; P00804.
DR PIR: P50329; P50329.
DR HSSP: P26663; 1UXP.
DR GO: GO:0016021; C:integral to membrane; IEA.

```

DR GO: GO:0019028; C:viral capsid; IEA.
 DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO: GO:0005489; F:electron transporter activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008226; F:serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F:structural molecule activity; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR GO: GO:0006118; F:electron transport; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; P:transcription; IEA.
 DR GO: GO:0019079; P:viral genome replication; IEA.
 DR GO: GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR003003; Cys_ser_1trypsin.
 DR InterPro: IPR000345; CysC_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR00745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002166; HCV_NS5a.
 DR InterPro: IPR001650; HCV RdRp.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR Pfam: PF01662; HCV_NS1; 1.
 DR Pfam: PF01662; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 DR Coat protein, Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 KW SEQUENCE 3010 AA; 327324 MW; 3DE6CF249BD151C CRC64;

Query Match 96.9%; Score 1485; DB 12; Length 3010;
 Best Local Similarity 95.5%; Pred. No. 2.3e-123;
 Matches 279; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGIIRACMVRKAGGVYQVAFMKALATGTYYVDHLPLODMAGLRDLAVAVEPY 60
 DB 915 AAGIIRACMVRKAGGVYQVAFMKALATGTYYVDHLPLODMAGLRDLAVAVEPY 974
 QY 61 IFSMVEKIIITWADTAACGDIISGLFVSARGRREIILGPADNPEGCGFLAPITAYSQ 120
 DB 975 VFSMDEKIIITWADTAACGDIISGLFVSARGRREIILGPADNPEGCGFLAPITAYSQ 1034
 QY 121 QTRLLGIIITSLGRKRVNVEGVVSTATOSFLATCVNGVCMVTFHAGSGKTLAAGP 180
 DB 1035 QTRLLGIIITSLGRKRVNVEGVVSTATOSFLATCVNGVCMVTFHAGSGKTLAAGP 1094
 QY 181 GPITQNTVNDQDLVGMQAPPGARSMTPTCGSSDLYLVTRADVTPVRRRGDSRGLLS 240
 DB 1095 GPITQNTVNDQDLVGMQAPPGARSMTPTCGSSDLYLVTRADVTPVRRRGDSRGLLS 1154
 QY 241 PRPVSYLKSGSGGPLLCPSGHVAIGIFRAAVCTRGVAADVPIFESMETMR 292
 DB 1155 PRPVSYLKSGSGGPLLCPSGHVAIGIFRAAVCTRGVAADVPIFESMETMR 1206

RESULT 13
 ID 0807P3 PRELIMINARY; PRT; 3010 AA.
 AC 0807P3;
 DT 01-JUN-2003 (TREMBLrel. 24, Created)
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCBI_TaxId=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MILE;
 RX MEDLINE:22047193; PubMed=12051758;
 RA Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,
 RA Nio Y., Hosaka M., Miyazaki Y., Shimotozono K.;
 RT "Subgenomic replicon derived from a cell line infected with the
 RT Hepatitis C virus."
 RL Biochem. Biophys. Res. Commun. 293:993-999(2002).
 DR EMBL: AB080299; BAC54896.1; -;
 DR GO: GO:0019028; C:viral capsid; IEA.
 DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO: GO:0005489; F:electron transporter activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008226; F:serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F:structural molecule activity; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR GO: GO:0006118; F:electron transport; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; P:transcription; IEA.
 DR GO: GO:0019079; P:viral genome replication; IEA.
 DR GO: GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR003003; Cys_ser_1trypsin.
 DR InterPro: IPR000345; CysC_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR00745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002166; HCV_NS5a.
 DR InterPro: IPR001650; HCV RdRp.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR Pfam: PF01662; HCV_NS1; 1.
 DR Pfam: PF01662; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 KW Polyprotein.
 SQ SEQUENCE 3010 AA; 327097 MW; EE6418C7A723B686 CRC64;

Query Match 96.9%; Score 1485; DB 12; Length 3010;
 Best Local Similarity 96.2%; Pred. No. 2.3e-123;

Matches 281; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLRACMLVKAAGGHVQMAFMKLAALTGTYYDHTLPLODMAHAGRLDAVAVEPV 60

DB 915 AAGLRACMLVKAAGGHVQMAFMKLAALTGTYYDHTLPLODMAHAGRLDAVAVEPV 974

QY 61 IFSDMEVKIITWGADTAACGDIISGLFVSARGREILLGPADNFEQGMRLAPITAYSQ 120

DB 975 VFSDMETKIIITWGADTAACGDIISGLFVSARGREILLGPADNFEQGMRLAPITAYSQ 1034

QY 121 QTRRLGCIITSLTGRDKNQVGEVQVSTATOSFLATCTNGVCMWTFHAGSKTLAGPK 180

DB 1035 QTRRLGCIITSLTGRDKNQVGEVQVSTATOSFLATCTNGVCMWTFHAGSKTLAGPK 1094

QY 181 GPITQMTYVDDLVGMQAPPGASMTPTCGSSDLVLTTRHADVIPRRRGDSRGLLS 240

DB 1095 GPITQMTYVDDLVGMQAPPGASMTPTCGSSDLVLTTRHADVIPRRRGDSRGLLS 1154

QY 241 PRPVSYLKSGSGGLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVSMETTM 292

DB 1155 PRPVSYLKSGSGGLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVSMETTM 1206

RESULT 14

Q9J311 PRELIMINARY; PRT; 3010 AA.

AC Q9J311

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

NC NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MD11.

RA Nagaya K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;

RT "Characteristics of hepatitis C viral genome associated with disease progression."

RL Submitted (NCV-1999) to the EMBL/Genbank/DBJ databases.

CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND RNA (BY SIMILARITY).

DR EMBL; AF207752; AAF65942.1; -.

DR PIR; A61196; A61196.

DR PIR; P00246; P00246.

DR PIR; PS0329; PS0329.

DR HSRP; P26663; LUXP.

DR GO; GO:0016031; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F:electron transporter activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008286; F:serine-type peptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0006118; P:electron transport; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P:transcription; IEA.

DR GO; GO:0019079; P:viral genome replication; IEA.

DR GO; GO:0019087; P:viral transformation; IEA.

DR InterPro; IPR009003; Cys_ser_lypsin.

DR InterPro; IPR000345; CytC_heme_BS.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV capsid.

DR InterPro; IPR002521; HCV core.

DR InterPro; IPR002519; HCV_env.

DR InterPro; IPR002511; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV NS5b.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR004109; Peptidase_C29.

DR InterPro; IPR007095; RNA_pol_D5_PS.

DR InterPro; IPR007094; RNA_pol_P5vir.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01539; HCV_core; 1.

DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; helicase_C; 1.

DR Pfam; PF00998; Viral_RdRp; 1.

DR Pfam; PD186062; HCV NS1; 1.

DR SMART; SM00487; DEXDC; 1.

DR PROSITE; PS00190; CYTOCHROME C; 1.

DR Coat protein; RNA-directed RNA polymerase; Glycoprotein; Nonstructural protein;

KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.

SQ SEQUENCE 3010 AA; 327193 MW; F88BA81174E19444 CRC64;

Query Match 96.7%; Score 1482; DB 12; length 3010;

Best Local Similarity 95.5%; Pred. No. 4.2e-123;

Matches 279; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLRACMLVKAAGGHVQMAFMKLAALTGTYYDHTLPLODMAHAGRLDAVAVEPV 60

DB 915 AAGLRACMLVKAAGGHVQMAFMKLAALTGTYYDHTLPLODMAHAGRLDAVAVEPV 974

QY 61 IFSDMEVKIITWGADTAACGDIISGLFVSARGREILLGPADNFEQGMRLAPITAYSQ 120

DB 975 VFSDMETKIIITWGADTAACGDIISGLFVSARGREILLGPADNFEQGMRLAPITAYSQ 1034

QY 121 QTRRLGCIITSLTGRDKNQVGEVQVSTATOSFLATCTNGVCMWTFHAGSKTLAGPK 180

DB 1035 QTRRLGCIITSLTGRDKNQVGEVQVSTATOSFLATCTNGVCMWTFHAGSKTLAGPK 1094

QY 181 GPITQMTYVDDLVGMQAPPGASMTPTCGSSDLVLTTRHADVIPRRRGDSRGLLS 240

DB 1095 GPITQMTYVDDLVGMQAPPGASMTPTCGSSDLVLTTRHADVIPRRRGDSRGLLS 1154

QY 241 PRPVSYLKSGSGGLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVSMETTM 292

DB 1155 PRPVSYLKSGSGGLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVSMETTM 1206

RESULT 15

O70822 PRELIMINARY; PRT; 361 AA.

AC O70822

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Polyprotein (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

NC NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89321154; PubMed=9656998;

RA Yamada K., Mori A., Seki W., Kimura J., Yuasa S., Matsuura Y.,

RT Miyamura T.;

RL "Critical point mutations for hepatitis C virus NS3 proteinase.";

RN Virology 246:104-112(1998).

[2]

RP SEQUENCE FROM N.A.
RA Mori A., Yamada K., Kimura J., Koide T., Yuasa S., Yamada E.,
"Enzymatic characterization of purified NS3 serine proteinase of
hepatitis C virus expressed in Escherichia coli.";
RL FEBS Lett. 378:37-42(1998).
DR EMBL: AB013627; BAA28505.1; -
DR HSP; P27958; 1A1V
DR GO: GO:0008236; P:serine-type peptidase activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO: GO:0019087; P:viral transformation; IEA.
DR InterPro: IPR009003; Cys_ser_trypsin.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; Peptidase_C29.
DR Pfam; PF01518; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 361 AA; 38144 MW; FE848F746A383FC CRC64;

Query Match 96.7%; Score 1481; DB 12; Length 361;
Best local similarity 95.2%; Pred. No. 3,1e-124;
Matches 278; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 1 AQLIRACMLVRKAAGHYVQMAFMKLAALTGYVDHLPLODMAHAGLRDIAVAPEV 60
DB 16 AQLIRACMLVRKAAGHYVQMAFMKLAALTGYVDHLPLODMAHAGLRDIAVAPEV 75

QY 61 IFSMEVKITWGDPTACGDITISGLPVSARGRKELIGPADSFEGQMRLLAPITAYSQ 120
DB 76 VFSAMETKVTWGDPTACGDITISGLPVSARGRKELIGPADSFEGQMRLLAPITAYSQ 135

QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVQWTFHGAGSKTLA GPK 180
DB 136 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVQWTFHGAGSKTLA GPK 195

QY 181 GPITQMTTNVDQDLVNGQAPPGASMTPTCTGSSDLVLTTRHADVLPYRRGDSRGSLLS 240
DB 196 GPITQMTTNVDQDLVNGQAPPGASMTPTCTGSSDLVLTTRHADVLPYRRGDSRGSLLS 255

QY 241 PRPVSYLKSGSGGFLCPSGHGVGIFRAVCTRGVAKAVDFIPVESMETMR 292
DB 256 PRPVSYLKSGSGGFLCPSGHGVGIFRAVCTRGVAKAVDFIPVESMETMR 307

Search completed: May 6, 2004, 09:35:50
Job time : 28.3102 secs

Fri May 7 13:37:06 2004

us-10-650-585-15.ra1

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 11.4605 Seconds
(without alignments)
1315.364 Million cell updates/sec

Title: US-10-650-585-15
Perfect score: 1532
Sequence: 1 AGLIRACMLVRRKAGGHV.....RGVAKAVDFIPVSMETMR 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/6D.COMB.pep:*

Fred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1473	96.1	1692	3	US-09-263-933-4	Sequence 4, Appl
2	1473	96.1	1692	4	US-09-919-901-4	Sequence 4, Appl
3	1473	96.1	2201	4	US-09-539-601-6	Sequence 6, Appl
4	1473	96.1	2201	4	US-09-539-601-15	Sequence 15, Appl
5	1473	96.1	2307	3	US-08-263-933-2	Sequence 2, Appl
6	1473	96.1	2307	4	US-09-919-901-2	Sequence 2, Appl
7	1473	96.1	3010	4	US-09-539-601-3	Sequence 3, Appl
8	1473	96.1	3010	4	US-09-539-601-21	Sequence 21, Appl
9	1473	96.1	3010	4	US-09-539-601-27	Sequence 27, Appl
10	1470	96.0	1692	3	US-09-263-933-11	Sequence 11, Appl
11	1470	96.0	2307	3	US-08-263-933-9	Sequence 9, Appl
12	1470	96.0	2307	4	US-09-919-901-9	Sequence 9, Appl
13	1466	95.7	1692	3	US-09-539-601-33	Sequence 18, Appl
14	1466	95.4	3010	4	US-09-263-933-18	Sequence 18, Appl
15	1461	95.4	1692	3	US-09-263-933-16	Sequence 16, Appl
16	1461	95.4	2307	3	US-09-919-901-18	Sequence 18, Appl
17	1461	95.4	2307	4	US-09-263-933-16	Sequence 16, Appl
18	1461	95.4	2307	4	US-09-919-901-16	Sequence 16, Appl
19	1450	94.6	3010	3	US-09-014-416-3	Sequence 3, Appl
20	1424	93.0	2013	1	US-08-324-977-12	Sequence 12, Appl
21	1424	93.0	2013	2	US-08-384-616-12	Sequence 12, Appl
22	1424	93.0	2013	2	US-08-904-686A-12	Sequence 12, Appl
23	1424	93.0	2013	3	US-09-315-850-12	Sequence 12, Appl
24	1424	93.0	2201	4	US-08-952-981A-2	Sequence 2, Appl
25	1424	93.0	2620	1	US-08-324-977-32	Sequence 32, Appl
26	1424	93.0	2620	1	US-08-384-616-32	Sequence 32, Appl
27	1424	93.0	2620	2	US-08-904-686A-32	Sequence 32, Appl

28	1424	93.0	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1424	93.0	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1424	93.0	2621	2	US-08-384-616-36	Sequence 36, Appl
31	1424	93.0	2621	2	US-08-904-686A-36	Sequence 36, Appl
32	1424	93.0	2621	3	US-09-315-850-36	Sequence 36, Appl
33	1424	93.0	3010	1	US-08-324-977-2	Sequence 2, Appl
34	1424	93.0	3010	1	US-08-324-977-14	Sequence 14, Appl
35	1424	93.0	3010	2	US-08-384-616-2	Sequence 2, Appl
36	1424	93.0	3010	2	US-08-384-616-14	Sequence 14, Appl
37	1424	93.0	3010	2	US-08-904-686A-2	Sequence 2, Appl
38	1424	93.0	3010	2	US-08-904-686A-14	Sequence 14, Appl
39	1424	93.0	3010	3	US-09-315-850-2	Sequence 2, Appl
40	1424	93.0	3010	3	US-09-315-850-14	Sequence 14, Appl
41	1362	88.9	3012	3	US-08-811-566-2	Sequence 2, Appl
42	1362	88.9	3012	4	US-09-034-756-2	Sequence 2, Appl
43	1359	86.7	1648	1	US-08-188-281B-12	Sequence 12, Appl
44	1359	86.7	1648	5	PCT-US94-07280-12	Sequence 12, Appl
45	1359	86.7	1648	5	PCT-US95-01087-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-263-933-4
Sequence 4, Application US/09263933
Patent No. 6280940
GENERAL INFORMATION:
APPLICANT: Potts, Karen B.
APPLICANT: Jackson, Roberta L.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/263,933
CURRENT FILING DATE: 1999-03-08
EARLIER APPLICATION NUMBER: 09/129,611
EARLIER FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1692
TYPE: PRT
ORGANISM: Artificial Sequence
US-09-263-933-4

Query Match 96.1%; Score 1473; DB 3; Length 1692;
Best Local Similarity 94.5%; Pred. No. 8.4e-139;
Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY	1	AGLIRACMLVRRKAGGHVQMAFMKLAALGTYYVDHLPQDNHAGRLAVAVEPV	60
DB	194	AGGLHACMLVRRKAGGHVQMAFMKLAALGTYYVDHLPQDNHAGRLAVAVEPV	253
QY	61	IFSDMEVKIITGADTAAAGDIIISGLPVARSRGREILIGPADNFGQGRLLAPITAVSQ	120
DB	254	VSDMEVKIITGADTAAAGDIIISGLPVARSRGREILIGPADNFGQGRLLAPITAVSQ	313
QY	121	QTRGLGCIITSLTGRDKNQVGEVQVSTAQSFATVNVGCVTFVFGASSKTLAGK	180
DB	314	QTRGLGCIITSLTGRDKNQVGEVQVSTAQSFATVNVGCVTFVFGASSKTLAGK	373
QY	181	GPITQWTVNDDDLVQWQAPPGARSMTPTCCSSDLVLTBRADVIPIVRRDSRGSILS	240
DB	374	GPITQWTVNDDDLVQWQAPPGARSMTPTCCSSDLVLTBRADVIPIVRRDSRGSILS	433
QY	241	PPRVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR	292
DB	434	PPRVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR	485

RESULT 2
US-09-919-901-4

Sequence 4, Application US/09919901
Patent No. 6599738
GENERAL INFORMATION:
APPLICANT: Potts, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/919,901
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 09/263,933
PRIOR FILING DATE: 1999-02-08
PRIOR APPLICATION NUMBER: 09/129,611
PRIOR FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1692
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: :
US-09-919-901-4

Query Match 96.1%; Score 1473; DB 4; Length 1692;
Best Local Similarity 94.5%; Pred. No. 8.4e-139;
Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLPLDQMAHAGLRDLAAVEPV 60
DB 194 AAGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLPLDQMAHAGLRDLAAVEPV 253
QY 61 IFSDMEVKIITWGADTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 120
DB 254 VFSDMETKVIITWGADTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 313
QY 121 QTRGLGCIITSLTGRDRNQVEGVYSTATQSFLLATCVNGVCMVTFHAGSKTLAGPK 180
DB 314 QTRGLGCIITSLTGRDRNQVEGVYSTATQSFLLATCVNGVCMVTFHAGSKTLAGPK 373
QY 181 GPITQMTNNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGRDSRGLS 240
DB 374 GPITQMTNNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGRDSRGLS 433
QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 292
DB 434 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 485

RESULT 3
US-09-539-601-6
Sequence 6, Application US/09539601C
Patent No. 6630343
GENERAL INFORMATION:
APPLICANT: Bartenschlager, Ralf FW
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
FILE REFERENCE: all sequences
CURRENT APPLICATION NUMBER: US/09/539,601C
CURRENT FILING DATE: 2001-08-30
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
NUMBER OF SEQ ID NOS: 51
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 2201
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-539-601-6

Query Match 96.1%; Score 1473; DB 4; Length 2201;
Best Local Similarity 94.5%; Pred. No. 1.2e-138;
Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLPLDQMAHAGLRDLAAVEPV 60
DB 106 AAGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLPLDQMAHAGLRDLAAVEPV 165
QY 61 IFSDMEVKIITWGADTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 120
DB 166 VFSDMETKVIITWGADTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 225
QY 121 QTRGLGCIITSLTGRDRNQVEGVYSTATQSFLLATCVNGVCMVTFHAGSKTLAGPK 180
DB 226 QTRGLGCIITSLTGRDRNQVEGVYSTATQSFLLATCVNGVCMVTFHAGSKTLAGPK 285
QY 181 GPITQMTNNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGRDSRGLS 240
DB 286 GPITQMTNNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGRDSRGLS 345
QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 292
DB 346 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 397

RESULT 4
US-09-539-601-15
Sequence 15, Application US/09539601C
Patent No. 6630343
GENERAL INFORMATION:
APPLICANT: Bartenschlager, Ralf FW
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
FILE REFERENCE: all sequences
CURRENT APPLICATION NUMBER: US/09/539,601C
CURRENT FILING DATE: 2001-08-30
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
NUMBER OF SEQ ID NOS: 51
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 15
LENGTH: 2201
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-539-601-15

Query Match 96.1%; Score 1473; DB 4; Length 2201;
Best Local Similarity 94.9%; Pred. No. 1.2e-138;
Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLPLDQMAHAGLRDLAAVEPV 60
DB 106 AAGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLPLDQMAHAGLRDLAAVEPV 165
QY 61 IFSDMEVKIITWGADTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 120
DB 166 VFSDMETKVIITWGADTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 225
QY 121 QTRGLGCIITSLTGRDRNQVEGVYSTATQSFLLATCVNGVCMVTFHAGSKTLAGPK 180
DB 226 QTRGLGCIITSLTGRDRNQVEGVYSTATQSFLLATCVNGVCMVTFHAGSKTLAGPK 285
QY 181 GPITQMTNNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGRDSRGLS 240
DB 286 GPITQMTNNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGRDSRGLS 345
QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 292
DB 346 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 397

RESULT 5
US-09-263-933-2
Sequence 2, Application US/09263933
Patent No. 6280940
GENERAL INFORMATION:
APPLICANT: Potts, Karen E.

APPLICANT: Jackson, Roberta L.
 APPLICANT: Patrick, Amy K.
 TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
 TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
 FILE REFERENCE: 0125-0005A
 CURRENT APPLICATION NUMBER: US/09/263,933
 CURRENT FILING DATE: 1999-03-08
 EARLIER APPLICATION NUMBER: 09/129,611
 EARLIER FILING DATE: 1998-08-05
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO: 2
 LENGTH: 2307
 TYPE: PRT
 ORGANISM: Artificial Sequence
 US-09-263-933-2

Query Match 96.1%; Score 1473; DB 3; Length 2307;
 Best Local Similarity 94.5%; Pred. No. 1,3e-138;
 Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLRACMLVRKAGGHVYQAFMKLAALGTYYVDHLTFLQDMAHAGLRDLAAVEPV 60
 DB 286 AAGLRACMLVRKAGGHVYQAFMKLAALGTYYVDHLTFLQDMAHAGLRDLAAVEPV 345
 QY 61 IFSDMEVKITTWGADTAACGDIISGLPVASRRREIILGPADNFEQGMRLAPITAYSQ 120
 DB 346 VPSDMETKITTWGADTAACGDIISGLPVASRRREIILGPADNFEQGMRLAPITAYSQ 405
 QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 180
 DB 406 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 465
 QY 181 GPITQMTNVDDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVRRGDSRGLLS 240
 DB 466 GPITQMTNVDDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVRRGDSRGLLS 525
 QY 241 PRPVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKADVPIVBSMETTKR 292
 DB 526 PRPVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKADVPIVBSMETTKR 577

RESULT 6
 US-09-919-901-2
 Sequence 2, Application US/09919901

GENERAL INFORMATION:
 APPLICANT: Potts, Karen E.
 APPLICANT: Jackson, Roberta L.
 APPLICANT: Patrick, Amy K.
 TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
 TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
 FILE REFERENCE: 0125-0005A
 CURRENT APPLICATION NUMBER: US/09/919,901
 CURRENT FILING DATE: 2001-08-02
 PRIOR APPLICATION NUMBER: 09/263,933
 PRIOR FILING DATE: 1999-02-08
 PRIOR APPLICATION NUMBER: 09/129,611
 PRIOR FILING DATE: 1998-08-05
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO: 2
 LENGTH: 2307
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION:
 US-09-919-901-2

Query Match 96.1%; Score 1473; DB 4; Length 2307;
 Best Local Similarity 94.5%; Pred. No. 1,3e-138;
 Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLRACMLVRKAGGHVYQAFMKLAALGTYYVDHLTFLQDMAHAGLRDLAAVEPV 60
 DB 286 AAGLRACMLVRKAGGHVYQAFMKLAALGTYYVDHLTFLQDMAHAGLRDLAAVEPV 345
 QY 61 IFSDMEVKITTWGADTAACGDIISGLPVASRRREIILGPADNFEQGMRLAPITAYSQ 120
 DB 346 VPSDMETKITTWGADTAACGDIISGLPVASRRREIILGPADNFEQGMRLAPITAYSQ 405
 QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 180
 DB 406 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 465
 QY 181 GPITQMTNVDDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVRRGDSRGLLS 240
 DB 466 GPITQMTNVDDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVRRGDSRGLLS 525
 QY 241 PRPVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKADVPIVBSMETTKR 292
 DB 526 PRPVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKADVPIVBSMETTKR 577

RESULT 7
 US-09-539-601-3
 Sequence 3, Application US/09539601C
 Patent No. 6630343
 GENERAL INFORMATION:
 APPLICANT: Bartsch, Ralf FM
 TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 FILE REFERENCE: all sequences
 CURRENT APPLICATION NUMBER: US/09/539,601C
 CURRENT FILING DATE: 2001-08-30
 EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
 EARLIER FILING DATE: 1999-04-03
 NUMBER OF SEQ ID NOS: 51
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO: 3
 LENGTH: 3010
 TYPE: PRT
 ORGANISM: Hepatitis C virus
 US-09-539-601-3

Query Match 96.1%; Score 1473; DB 4; Length 3010;
 Best Local Similarity 94.9%; Pred. No. 1,9e-138;
 Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLRACMLVRKAGGHVYQAFMKLAALGTYYVDHLTFLQDMAHAGLRDLAAVEPV 60
 DB 915 AAGLRACMLVRKAGGHVYQAFMKLAALGTYYVDHLTFLQDMAHAGLRDLAAVEPV 974
 QY 915 AAGLRACMLVRKAGGHVYQAFMKLAALGTYYVDHLTFLQDMAHAGLRDLAAVEPV 974
 DB 915 AAGLRACMLVRKAGGHVYQAFMKLAALGTYYVDHLTFLQDMAHAGLRDLAAVEPV 974
 QY 61 IFSDMEVKITTWGADTAACGDIISGLPVASRRREIILGPADNFEQGMRLAPITAYSQ 120
 DB 975 VPSDMETKITTWGADTAACGDIISGLPVASRRREIILGPADNFEQGMRLAPITAYSQ 1034
 QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 180
 DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 1094
 QY 181 GPITQMTNVDDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVRRGDSRGLLS 240
 DB 1095 GPITQMTNVDDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVRRGDSRGLLS 1154
 QY 241 PRPVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKADVPIVBSMETTKR 292
 DB 1155 PRPVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKADVPIVBSMETTKR 1206

RESULT 8
 US-09-539-601-21
 Sequence 21, Application US/09539601C
 Patent No. 6630343
 GENERAL INFORMATION:
 APPLICANT: Bartsch, Ralf FM
 TITLE OF INVENTION: Hepatitis C Virus Cell Culture System

FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 3010
; TYPE: PR
; ORGANISM: Hepatitis C virus
US-09-539-601-21

Query Match 96.1%; Score 1473; DB 4; Length 3010;
Best Local Similarity 94.9%; Pred. No. 1.9e-138;
Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 60
DB 915 AHGILIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 974
QY 61 IFSMEVKITITWADTACGDIISGLPVSARRGREILGPADNFEQGMRLAPITAYSQ 120
DB 975 VFSDEMKVITWADTACGDIISGLPVSARRGREILGPADNFEQGMRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 180
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 1094
QY 181 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVTPVRRGDSRGLLS 240
DB 1095 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVTPVRRGDSRGLLS 1154
QY 241 PRPVSYLKSGSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 292
DB 1155 PRPVSYLKSGSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 1206

RESULT 9
US-09-539-601-27
; Sequence 27, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Barenscnager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 3010
; TYPE: PR
; ORGANISM: Hepatitis C virus
US-09-539-601-27

Query Match 96.1%; Score 1473; DB 4; Length 3010;
Best Local Similarity 94.9%; Pred. No. 1.9e-138;
Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 60
DB 915 AHGILIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 974
QY 61 IFSMEVKITITWADTACGDIISGLPVSARRGREILGPADNFEQGMRLAPITAYSQ 120
DB 975 VFSDEMKVITWADTACGDIISGLPVSARRGREILGPADNFEQGMRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 180
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 1094

QY 181 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVTPVRRGDSRGLLS 240
DB 1095 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVTPVRRGDSRGLLS 1154
QY 241 PRPVSYLKSGSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 292
DB 1155 PRPVSYLKSGSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 1206

RESULT 10
US-09-263-933-11
; Sequence 11, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potte, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PR
; ORGANISM: Artificial Sequence
US-09-263-933-11

Query Match 96.0%; Score 1470; DB 3; Length 1692;
Best Local Similarity 94.2%; Pred. No. 1.7e-118;
Matches 275; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 60
DB 194 AAGLIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 253
QY 61 IFSMEVKITITWADTACGDIISGLPVSARRGREILGPADNFEQGMRLAPITAYSQ 120
DB 254 VFSDEMKVITWADTACGDIISGLPVSARRGREILGPADNFEQGMRLAPITAYSQ 313
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 180
DB 314 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 373
QY 181 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVTPVRRGDSRGLLS 240
DB 374 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVTPVRRGDSRGLLS 433
QY 241 PRPVSYLKSGSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 292
DB 434 PRPVSYLKSGSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 485

RESULT 11
US-09-919-901-11
; Sequence 11, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potte, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08

QY 1 AAGLIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 60
DB 915 AHGILIRACMLVRKAGAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 974
QY 61 IFSMEVKITITWADTACGDIISGLPVSARRGREILGPADNFEQGMRLAPITAYSQ 120
DB 975 VFSDEMKVITWADTACGDIISGLPVSARRGREILGPADNFEQGMRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 180
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 1094

PRIOR APPLICATION NUMBER: 09/129,611
 PRIOR FILING DATE: 1998-08-05
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO 11
 LENGTH: 1692
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: :
 US-09-919-901-11

Query Match 96.0%; Score 1470; DB 4; Length 1692;
 Best Local Similarity 94.2%; Pred. No. 1,76-138;
 Matches 275; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

1 AAGLIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAVEPV 60
 194 AAGLIHACMLVRKAGHYVQMAFMKLGALTGTYYNHLPLRDMAHAGRLDAVAVEPV 253
 61 IFSDEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGMRLAPITAYSQ 120
 254 VFSDEMKIITWGADTAACGDIILGLPVASARGREIILGPADSLBGRGMRLLAPITAYSQ 313
 121 QTRGLGCIITSLTRGRDNQVGEVQVSTATOSFLATCVNGVCWTFHAGSKTLAAGPK 180
 314 QTRGLGCIITSLTRGRDNQVGEVQVSTATOSFLATCVNGVCWTFHAGSKTLAAGPK 373
 181 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGLLS 240
 374 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGLLS 433
 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFPVESMETTMR 292
 434 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFPVESMETTMR 485

RESULT 12

US-09-263-933-9
 Sequence 9, Application US/09263933
 Patent No. 6280940
 GENERAL INFORMATION:
 APPLICANT: Potts, Karen E.
 APPLICANT: Jackson, Roberta L.
 APPLICANT: Patrick, Amy K.
 TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
 TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
 FILE REFERENCE: 0125-0005A
 CURRENT APPLICATION NUMBER: US/09/263,933
 CURRENT FILING DATE: 1999-03-08
 EARLIER APPLICATION NUMBER: 09/129,611
 EARLIER FILING DATE: 1998-08-05
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO 9
 LENGTH: 2307
 TYPE: PRT
 ORGANISM: Artificial Sequence
 US-09-263-933-9

Query Match 96.0%; Score 1470; DB 3; Length 2307;
 Best Local Similarity 94.2%; Pred. No. 2,66-138;
 Matches 275; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

1 AAGLIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAVEPV 60
 286 AAGLIHACMLVRKAGHYVQMAFMKLGALTGTYYNHLPLRDMAHAGRLDAVAVEPV 345
 61 IFSDEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGMRLAPITAYSQ 120
 346 VFSDEMKIITWGADTAACGDIILGLPVASARGREIILGPADSLBGRGMRLLAPITAYSQ 405
 121 QTRGLGCIITSLTRGRDNQVGEVQVSTATOSFLATCVNGVCWTFHAGSKTLAAGPK 180

406 QTRGLGCIITSLTRGRDNQVGEVQVSTATOSFLATCVNGVCWTFHAGSKTLAAGPK 465
 181 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGLLS 240
 466 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGLLS 525
 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFPVESMETTMR 292
 526 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFPVESMETTMR 577

RESULT 13
 US-09-919-901-9
 Sequence 9, Application US/09919901
 Patent No. 6539738
 GENERAL INFORMATION:
 APPLICANT: Potts, Karen E.
 APPLICANT: Jackson, Roberta L.
 APPLICANT: Patrick, Amy K.
 TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
 TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
 FILE REFERENCE: 0125-0005A
 CURRENT APPLICATION NUMBER: US/09/919,901
 CURRENT FILING DATE: 2001-08-02
 PRIOR FILING DATE: 1999-02-08
 PRIOR APPLICATION NUMBER: 09/263,933
 PRIOR FILING DATE: 1998-08-05
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO 9
 LENGTH: 2307
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: :
 US-09-919-901-9

Query Match 96.0%; Score 1470; DB 4; Length 2307;
 Best Local Similarity 94.2%; Pred. No. 2,66-138;
 Matches 275; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

1 AAGLIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAVEPV 60
 286 AAGLIHACMLVRKAGHYVQMAFMKLGALTGTYYNHLPLRDMAHAGRLDAVAVEPV 345
 61 IFSDEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGMRLAPITAYSQ 120
 346 VFSDEMKIITWGADTAACGDIILGLPVASARGREIILGPADSLBGRGMRLLAPITAYSQ 405
 121 QTRGLGCIITSLTRGRDNQVGEVQVSTATOSFLATCVNGVCWTFHAGSKTLAAGPK 180
 406 QTRGLGCIITSLTRGRDNQVGEVQVSTATOSFLATCVNGVCWTFHAGSKTLAAGPK 465
 181 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGLLS 240
 466 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGLLS 525
 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFPVESMETTMR 292
 526 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFPVESMETTMR 577

RESULT 14
 US-09-539-601-33
 Sequence 33, Application US/09539601C
 Patent No. 6630343
 GENERAL INFORMATION:
 APPLICANT: Bartenschlager, Ralf FM
 TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 FILE REFERENCE: all sequences
 CURRENT APPLICATION NUMBER: US/09/539,601C.

CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-33

Query Match 95.7%; Score 1466; DB 4; Length 3010;
Best Local Similarity 94.5%; Pred. No. 9,7e-138;
Matches 276; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 1 ACGIIRACMLVRKAGGHYQMAFMKLAALTGYVDHLTPLODRAHAGIRDIAVAVEPV 60
DB 915 AHGILIRACMLVRKAGGHYQMAFMKLAALTGYVDHLTPLODRAHAGIRDIAVAVEPV 974
QY 61 IFSDMETKVTWAGADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 120
DB 975 VFSDMETKVTWAGADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHGAGSKTLAGPK 180
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHGAGSKTLAGPK 1094
QY 181 GPITQMTNNVDQDLYGMOAPPGARSMTPTCTGSSDLYLVTRHADVLPVRRRGRSGSLLS 240
DB 1095 GPITQMTNNVDQDLYGMOAPPGARSMTPTCTGSSDLYLVTRHADVLPVRRRGRSGSLLS 1154
QY 241 PRPVSYLKSGSGGFLCPSGHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 292
DB 1155 PRPVSYLKSGSGGFLCPSGHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 1206

RESULT 15
US-09-263-933-18
; Sequence 18, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potes, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; EARLIER FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-18

Query Match 95.4%; Score 1461; DB 3; Length 1692;
Best Local Similarity 93.8%; Pred. No. 1.4e-137;
Matches 274; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

QY 1 ACGIIRACMLVRKAGGHYQMAFMKLAALTGYVDHLTPLODRAHAGIRDIAVAVEPV 60
DB 194 ACGIIRACMLVRKAGGHYQMAFMKLAALTGYVDHLTPLODRAHAGIRDIAVAVEPV 253
QY 61 IFSDMETKVTWAGADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 120
DB 254 VFSDMETKVTWAGADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 313
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHGAGSKTLAGPK 180

DB 314 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHGAGSKTLAGPK 373
QY 181 GPITQMTNNVDQDLYGMOAPPGARSMTPTCTGSSDLYLVTRHADVLPVRRRGRSGSLLS 240
DB 374 GPITQMTNNVDQDLYGMOAPPGARSMTPTCTGSSDLYLVTRHADVLPVRRRGRSGSLLS 433
QY 241 PRPVSYLKSGSGGFLCPSGHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 292
DB 434 PRPVSYLKSGSGGFLCPSGHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 485

Search completed: May 6, 2004, 09:39:06
Job time: 12.4605 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 29.8706 Seconds

(without alignments)
2713.357 Million cell updates/sec

Title: US-10-650-585-15

Perfect score: 1532

Sequence: 1 AAGLIRACMLVRKAGSHV.....RGVAKAVDPFVNSMETTW 292

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 27566755 residues

1140673

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubppaa/PCR_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubppaa/PCRUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubppaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubppaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubppaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1532	100.0	292	13	US-10-017-736-15
2	1532	100.0	292	16	US-10-650-585-15
3	1532	100.0	303	13	US-10-017-736-10
4	1532	100.0	303	16	US-10-650-585-10
5	1532	100.0	334	13	US-10-017-736-4
6	1532	100.0	334	16	US-10-650-585-4
7	1532	100.0	341	13	US-10-017-736-14
8	1532	100.0	341	16	US-10-650-585-14
9	1532	100.0	352	13	US-10-017-736-13
10	1532	100.0	352	16	US-10-650-585-13
11	1532	100.0	380	13	US-10-017-736-12
12	1532	100.0	380	16	US-10-650-585-12
13	1532	100.0	393	13	US-10-017-736-11
14	1532	100.0	393	16	US-10-650-585-11
15	1532	100.0	409	13	US-10-017-736-2

16	1532	100.0	409	16	US-10-650-585-2	Sequence 2, Appli
17	1532	99.4	303	13	US-10-017-736-18	Sequence 18, Appli
18	1532	99.4	303	16	US-10-650-585-18	Sequence 18, Appli
19	1532	99.3	303	13	US-10-017-736-16	Sequence 16, Appli
20	1532	99.3	303	16	US-10-650-585-16	Sequence 16, Appli
21	1532	98.8	301	13	US-10-017-736-17	Sequence 17, Appli
22	1532	98.8	301	16	US-10-650-585-17	Sequence 17, Appli
23	1473	96.1	1992	10	US-09-919-901-4	Sequence 4, Appli
24	1473	96.1	1992	14	US-10-191-966-4	Sequence 4, Appli
25	1473	96.1	2201	13	US-10-029-907-3	Sequence 3, Appli
26	1473	96.1	2201	14	US-10-309-561-3	Sequence 3, Appli
27	1473	96.1	2307	10	US-09-919-901-2	Sequence 2, Appli
28	1473	96.1	2307	14	US-10-191-966-2	Sequence 2, Appli
29	1473	96.1	3010	12	US-10-467-000-1	Sequence 1, Appli
30	1470	96.0	1692	10	US-09-919-901-11	Sequence 11, Appli
31	1470	96.0	1692	14	US-10-191-966-11	Sequence 11, Appli
32	1470	96.0	2307	10	US-09-919-901-9	Sequence 9, Appli
33	1470	96.0	2307	14	US-10-191-966-9	Sequence 9, Appli
34	1461	95.4	1692	10	US-09-919-901-18	Sequence 18, Appli
35	1461	95.4	1692	14	US-10-191-966-18	Sequence 18, Appli
36	1461	95.4	2307	10	US-09-919-901-16	Sequence 16, Appli
37	1461	95.4	2307	14	US-10-191-966-16	Sequence 16, Appli
38	1461	93.0	2201	13	US-10-085-476-2	Sequence 2, Appli
39	1362	88.9	3011	9	US-09-742-659-4	Sequence 4, Appli
40	1362	88.9	3011	10	US-09-891-894-3	Sequence 3, Appli
41	1362	88.9	3011	15	US-10-184-150-3	Sequence 3, Appli
42	1362	88.9	3011	14	US-10-328-997-3	Sequence 2, Appli
43	1362	88.9	3012	9	US-09-238-076-2	Sequence 2, Appli
44	1362	88.9	3012	10	US-09-995-937-2	Sequence 2, Appli
45	1362	88.9	3012	10	US-09-917-563-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-017-736-15
; Sequence 15, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 292
; TYPE: PRT
; ORGANSIM: HCV
; US-10-017-736-15

QY	1	AAGLIRACMLVRKAGSHVYVMAFMKLAALGTYYYDHLTPQDMAHAGLRDLAAVEEV	60
DB	1	AAGLIRACMLVRKAGSHVYVMAFMKLAALGTYYYDHLTPQDMAHAGLRDLAAVEEV	60
QY	61	IFSDMEVKIITVAGDTAACGDIISGLPVASARRGREITLLGPADNFFGQGRLLIAPITVYSQ	120
DB	61	IFSDMEVKIITVAGDTAACGDIISGLPVASARRGREITLLGPADNFFGQGRLLIAPITVYSQ	120
QY	121	QTRGLGCIITLTGDRKQVGEVQVYVTAOSFLATVGVNVCMTVFHGAASKTLAPGX	180
DB	121	QTRGLGCIITLTGDRKQVGEVQVYVTAOSFLATVGVNVCMTVFHGAASKTLAPGX	180
QY	181	GPITQWYTNVDDILVGMQAPPGARSMTPTCCSSDLVYVTRHADYIPVRRDSSGSLIS	240
DB	181	GPITQWYTNVDDILVGMQAPPGARSMTPTCCSSDLVYVTRHADYIPVRRDSSGSLIS	240

Db 181 GPTQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVTPVRRRGRSGSLLS 240

QY 241 PRPVSYLKSSGGGGLTSPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292

Db 241 PRPVSYLKSSGGGGLTSPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292

RESULT 2

US-10-650-585-15

Sequence 15, Application US/10650585

Publication No. US20040077066A1

GENERAL INFORMATION:

APPLICANT: Boehringer Ingelheim (Canada) Ltd.

TITLE OF INVENTION: Purified Active HCV NS2/3 Protease

FILE REFERENCE: 13/082

CURRENT APPLICATION NUMBER: US/10/650,585

CURRENT FILING DATE: 2003-08-28

PRIOR APPLICATION NUMBER: US/10/017,736A

PRIOR FILING DATE: 2001-12-14

PRIOR APPLICATION NUMBER: 60/256,031

PRIOR FILING DATE: 2000-12-15

NUMBER OF SEQ ID NOS: 21

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 15

LENGTH: 292

TYPE: PR

ORGANISM: HCV

US-10-650-585-15

Query Match 100.0%; Score 1532; DB 16; Length 292;

Best Local Similarity 100.0%; Pred. No. 1e-148;

Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 60

Db 1 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 60

QY 61 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLAIPITAYSQ 120

Db 61 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLAIPITAYSQ 120

QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFPLATCNVGVMTVPHGAGSKTLAGPK 180

Db 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFPLATCNVGVMTVPHGAGSKTLAGPK 180

QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFPLATCNVGVMTVPHGAGSKTLAGPK 180

Db 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFPLATCNVGVMTVPHGAGSKTLAGPK 180

QY 181 GPTQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVTPVRRRGRSGSLLS 240

Db 181 GPTQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVTPVRRRGRSGSLLS 240

QY 241 PRPVSYLKSSGGGGLTSPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292

Db 241 PRPVSYLKSSGGGGLTSPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292

RESULT 3

US-10-017-736-10

Sequence 10, Application US/10017736

Publication No. US20020192640A1

GENERAL INFORMATION:

APPLICANT: Boehringer Ingelheim (Canada) Ltd.

TITLE OF INVENTION: Purified Active HCV NS2/3 Protease

FILE REFERENCE: 13/082

CURRENT APPLICATION NUMBER: US/10/017,736

CURRENT FILING DATE: 2001-12-14

PRIOR APPLICATION NUMBER: 60/256,031

PRIOR FILING DATE: 2000-12-15

NUMBER OF SEQ ID NOS: 21

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10

LENGTH: 303

TYPE: PR

ORGANISM: HCV

US-10-017-736-10

Query Match 100.0%; Score 1532; DB 13; Length 303;

Best Local Similarity 100.0%; Pred. No. 1.1e-148;

Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 60

Db 1 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 71

QY 61 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLAIPITAYSQ 120

Db 61 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLAIPITAYSQ 131

QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFPLATCNVGVMTVPHGAGSKTLAGPK 180

Db 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFPLATCNVGVMTVPHGAGSKTLAGPK 191

QY 181 GPTQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVTPVRRRGRSGSLLS 240

Db 181 GPTQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVTPVRRRGRSGSLLS 251

QY 241 PRPVSYLKSSGGGGLTSPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292

Db 241 PRPVSYLKSSGGGGLTSPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 303

RESULT 4

US-10-650-585-10

Sequence 10, Application US/10650585

Publication No. US20040077066A1

GENERAL INFORMATION:

APPLICANT: Boehringer Ingelheim (Canada) Ltd.

TITLE OF INVENTION: Purified Active HCV NS2/3 Protease

FILE REFERENCE: 13/082

CURRENT APPLICATION NUMBER: US/10/650,585

CURRENT FILING DATE: 2003-08-28

PRIOR APPLICATION NUMBER: US/10/017,736A

PRIOR FILING DATE: 2001-12-14

PRIOR APPLICATION NUMBER: 60/256,031

PRIOR FILING DATE: 2000-12-15

NUMBER OF SEQ ID NOS: 21

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10

LENGTH: 303

TYPE: PR

ORGANISM: HCV

US-10-650-585-10

Query Match 100.0%; Score 1532; DB 16; Length 303;

Best Local Similarity 100.0%; Pred. No. 1.1e-148;

Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 60

Db 1 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 71

QY 61 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLAIPITAYSQ 120

Db 61 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLAIPITAYSQ 131

QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFPLATCNVGVMTVPHGAGSKTLAGPK 180

Db 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFPLATCNVGVMTVPHGAGSKTLAGPK 191

QY 181 GPTQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVTPVRRRGRSGSLLS 240

Db 181 GPTQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVTPVRRRGRSGSLLS 251

QY 241 PRPVSYLKSSGGGGLTSPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292

Db 241 PRPVSYLKSSGGGGLTSPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 303

RESULT 5
US-10-017-736-4
; Sequence 4, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 334
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-4

Query Match 100.0%; Score 1532; DB 13; Length 334;
Best Local Similarity 100.0%; Pred. No. 1.2e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHVQVMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 60
DB 27 AAGLIRACMLVRKAGGHVQVMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 86
QY 61 IFSDMVEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 120
DB 87 IFSDMVEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 146
QY 121 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSFLATCVNGVCWTFVHGAGSKTLAAGPK 180
DB 147 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSFLATCVNGVCWTFVHGAGSKTLAAGPK 206
QY 181 GPITQWYTNVDODLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
DB 207 GPITQWYTNVDODLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 266
QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292
DB 267 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 318

RESULT 6
US-10-650-585-4
; Sequence 4, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 334
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-4

Query Match 100.0%; Score 1532; DB 16; Length 334;
Best Local Similarity 100.0%; Pred. No. 1.2e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHVQVMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 60
|||||

DB 27 AAGLIRACMLVRKAGGHVQVMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 86
QY 61 IFSDMVEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 120
DB 87 IFSDMVEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 146
QY 121 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSFLATCVNGVCWTFVHGAGSKTLAAGPK 180
DB 147 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSFLATCVNGVCWTFVHGAGSKTLAAGPK 206
QY 181 GPITQWYTNVDODLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
DB 207 GPITQWYTNVDODLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 266
QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292
DB 267 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 318

RESULT 7
US-10-017-736-14
; Sequence 14, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-14

Query Match 100.0%; Score 1532; DB 13; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.3e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHVQVMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 60
DB 50 AAGLIRACMLVRKAGGHVQVMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 109
QY 61 IFSDMVEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 120
DB 110 IFSDMVEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 169
QY 121 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSFLATCVNGVCWTFVHGAGSKTLAAGPK 180
DB 170 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSFLATCVNGVCWTFVHGAGSKTLAAGPK 229
QY 181 GPITQWYTNVDODLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
DB 230 GPITQWYTNVDODLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 289
QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292
DB 290 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 341

RESULT 8
US-10-650-585-14
; Sequence 14, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585

; CURRENT FILING DATE: 2003-08-28
 ; PRIOR APPLICATION NUMBER: US/10/017,736A
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14
 ; LENGTH: 341
 ; TYPE: PRT
 ; ORGANISM: HCV
 ; US-10-650-585-14

Query Match 100.0%; Score 1532; DB 16; Length 341;
 Best Local Similarity 100.0%; Pred. No. 1.3e-148;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGLRDLAVAVEPV 60
 DB 50 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGLRDLAVAVEPV 109
 QY 61 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLLAPITAYSQ 120
 DB 110 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLLAPITAYSQ 169
 QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 180
 DB 170 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 229
 QY 181 GPITOMTNTVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
 DB 230 GPITOMTNTVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 289
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292
 DB 290 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 341

RESULT 9

US-10-017-736-13
 ; Sequence 13, Application US/10017736
 ; Publication No. US20020192640A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/017,736
 ; CURRENT FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 352
 ; TYPE: PRT
 ; ORGANISM: HCV
 ; US-10-017-736-13

Query Match 100.0%; Score 1532; DB 13; Length 352;
 Best Local Similarity 100.0%; Pred. No. 1.3e-148;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGLRDLAVAVEPV 60
 DB 61 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGLRDLAVAVEPV 120
 QY 61 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLLAPITAYSQ 120
 DB 121 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLLAPITAYSQ 180
 QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 180
 DB 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 240

QY 181 GPITOMTNTVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
 DB 241 GPITOMTNTVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292
 DB 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 352

RESULT 10

US-10-650-585-13
 ; Sequence 13, Application US/10650585
 ; Publication No. US2004007066A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/650,585
 ; CURRENT FILING DATE: 2003-08-28
 ; PRIOR APPLICATION NUMBER: US/10/017,736A
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 352
 ; TYPE: PRT
 ; ORGANISM: HCV
 ; US-10-650-585-13

Query Match 100.0%; Score 1532; DB 16; Length 352;
 Best Local Similarity 100.0%; Pred. No. 1.3e-148;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGLRDLAVAVEPV 60
 DB 61 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGLRDLAVAVEPV 120
 QY 61 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLLAPITAYSQ 120
 DB 121 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLLAPITAYSQ 180
 QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 180
 DB 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 240
 QY 181 GPITOMTNTVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
 DB 241 GPITOMTNTVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292
 DB 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 352

RESULT 11

US-10-017-736-12
 ; Sequence 12, Application US/10017736
 ; Publication No. US20020192640A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/017,736
 ; CURRENT FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12
 ; LENGTH: 380

QY 1 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGLRDLAVAVEPV 60
 DB 61 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGLRDLAVAVEPV 120
 QY 61 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLLAPITAYSQ 120
 DB 121 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLLAPITAYSQ 180
 QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 180
 DB 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 240

```
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-12

Query Match      100.0%; Score 1532; DB 13; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.5e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AAGIIRACMLVRKAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAEVP 60
Db 89 AAGIIRACMLVRKAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAEVP 148
Cy 61 IFSDMVEKIIITWGAADTAACGDIISGLPVARSAREILLGPADNFEQGMRLAPITAYSQ 120
Db 149 IFSDMVEKIIITWGAADTAACGDIISGLPVARSAREILLGPADNFEQGMRLAPITAYSQ 208
Cy 121 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 180
Db 209 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 268
Cy 181 GPITQMTNVDDIVGMOAPRGASMTPTCGSSDLYLTRHADVIIVRRRGSRSGLLS 240
Db 269 GPITQMTNVDDIVGMOAPRGASMTPTCGSSDLYLTRHADVIIVRRRGSRSGLLS 328
Cy 241 PRPVSYLKSGSGGPLLCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 292
Db 329 PRPVSYLKSGSGGPLLCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 380

RESULT 12
; Sequence 12, Application US/10650585
; Publication No. US2004007066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 380
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-12

Query Match      100.0%; Score 1532; DB 16; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.5e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AAGIIRACMLVRKAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAEVP 60
Db 89 AAGIIRACMLVRKAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAEVP 148
Cy 61 IFSDMVEKIIITWGAADTAACGDIISGLPVARSAREILLGPADNFEQGMRLAPITAYSQ 120
Db 149 IFSDMVEKIIITWGAADTAACGDIISGLPVARSAREILLGPADNFEQGMRLAPITAYSQ 208
Cy 121 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 180
Db 209 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 268
Cy 181 GPITQMTNVDDIVGMOAPRGASMTPTCGSSDLYLTRHADVIIVRRRGSRSGLLS 240
Db 269 GPITQMTNVDDIVGMOAPRGASMTPTCGSSDLYLTRHADVIIVRRRGSRSGLLS 328
Cy 241 PRPVSYLKSGSGGPLLCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 292
Db 329 PRPVSYLKSGSGGPLLCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 292
```

```
Db 329 PRPVSYLKSGSGGPLLCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 380

RESULT 13
US-10-017-736-11
; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-11

Query Match      100.0%; Score 1532; DB 13; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.5e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AAGIIRACMLVRKAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAEVP 60
Db 102 AAGIIRACMLVRKAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAEVP 161
Cy 61 IFSDMVEKIIITWGAADTAACGDIISGLPVARSAREILLGPADNFEQGMRLAPITAYSQ 120
Db 162 IFSDMVEKIIITWGAADTAACGDIISGLPVARSAREILLGPADNFEQGMRLAPITAYSQ 221
Cy 121 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 180
Db 222 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 281
Cy 181 GPITQMTNVDDIVGMOAPRGASMTPTCGSSDLYLTRHADVIIVRRRGSRSGLLS 240
Db 282 GPITQMTNVDDIVGMOAPRGASMTPTCGSSDLYLTRHADVIIVRRRGSRSGLLS 341
Cy 241 PRPVSYLKSGSGGPLLCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 292
Db 342 PRPVSYLKSGSGGPLLCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 393

RESULT 14
US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US2004007066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-11

Query Match      100.0%; Score 1532; DB 16; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.5e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 AOGILIRACMLVRKAAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGLRDLAVAVEPV 60
DB      102 AOGILIRACMLVRKAAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGLRDLAVAVEPV 161
QY      61 IFSDMEVKIITWGDITACGDIISGLPVASARGREILGPADNFEQGWRLAPITAYSQ 120
DB      162 IFSDMEVKIITWGDITACGDIISGLPVASARGREILGPADNFEQGWRLAPITAYSQ 221
QY      121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 180
DB      222 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 281
QY      181 GPITOMTYNVQDQIVGQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
DB      282 GPITOMTYNVQDQIVGQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 341
QY      241 PRPVSYLKSSGGGFLCPSGHAVGIFRAAVCTRGVAKAVDFIPIVESMETTMR 292
DB      342 PRPVSYLKSSGGGFLCPSGHAVGIFRAAVCTRGVAKAVDFIPIVESMETTMR 393
```

RESULT 15

```
US-10-017-736-2
; Sequence 2, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017, 736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-2
```

```
Query Match      100.0%; Score 1532; DB 13; Length 409;
Best Local Similarity 100.0%; Pred. No. 1.6e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 AOGILIRACMLVRKAAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGLRDLAVAVEPV 60
DB      106 AOGILIRACMLVRKAAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGLRDLAVAVEPV 165
QY      61 IFSDMEVKIITWGDITACGDIISGLPVASARGREILGPADNFEQGWRLAPITAYSQ 120
DB      166 IFSDMEVKIITWGDITACGDIISGLPVASARGREILGPADNFEQGWRLAPITAYSQ 225
QY      121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 180
DB      226 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 285
QY      181 GPITOMTYNVQDQIVGQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
DB      286 GPITOMTYNVQDQIVGQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 345
QY      241 PRPVSYLKSSGGGFLCPSGHAVGIFRAAVCTRGVAKAVDFIPIVESMETTMR 292
DB      346 PRPVSYLKSSGGGFLCPSGHAVGIFRAAVCTRGVAKAVDFIPIVESMETTMR 397
```

Search completed: May 6, 2004, 09:43:23
Job time : 30.8706 secs